

92-371

Sampling and Weighting

1996 Census Technical Reports



3 1761 11631856 9

GOV
DOC.

CENSUS

Data in many forms

Statistics Canada disseminates data in a variety of forms. In addition to publications, both standard and special tabulations are offered. Data are available on the Internet, compact disc, diskette, computer printouts, microfiche and microfilm, and magnetic tape. Maps and other geographic reference materials are available for some types of data. Direct online access to aggregated information is possible through CANSIM, Statistics Canada's machine-readable database and retrieval system.

How to obtain more information

Inquiries about this product and related statistics or services should be directed to the Statistics Canada Regional Reference Centre in:

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(780) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

You can also visit our World Wide Web site: <http://www.statcan.ca>

Toll-free access is provided for all users who reside outside the local dialing area of any of the Regional Reference Centres.

National enquiries line

1 800 263-1136

National telecommunications device for the hearing impaired

1 800 363-7629

Order-only line (Canada and United States)

1 800 267-6677

Fax order line (Canada and United States)

1 877 287-4369

Ordering/Subscription information

All prices exclude sales tax

Catalogue No. 92-371-XPB is available for \$15.00 per copy in Canada. Outside Canada the cost is US \$15.00 per copy.

Please order by mail, at Statistics Canada, Dissemination Division, Circulation Management, 120 Parkdale Avenue, Ottawa, Ontario, K1A 0T6; by phone, at (613) 951-7277 or 1 800 700-1033; by fax, at (613) 951-1584 or 1 800 889-9734; or by Internet, at order@statcan.ca. For changes of address, please provide both old and new addresses. Statistics Canada products may also be purchased from authorized agents, bookstores and local Statistics Canada offices.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact your nearest Statistics Canada Regional Reference Centre.

ISBN 0-660-60571-6



9 780660 605715

92-371-XPB96000



Statistics Canada

Sampling and Weighting

1996 Census Technical Reports

Published by authority of the Minister
responsible for Statistics Canada

© Minister of Industry, 1999

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission from Licence Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

December, 1999

Catalogue No. 92-371-XPB

ISBN 0-660-60571-6

Ottawa

Note of Appreciation

Canada owes the success of its statistical system to a long-standing cooperation involving Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Canadian Cataloguing in Publication Data

Main entry under title:

Sampling and Weighting

(1996 Census technical reports)

Text in English and French on inverted pages.

Title on added t. p. title: Échantillonnage et pondération.

ISBN 0-660-60571-6

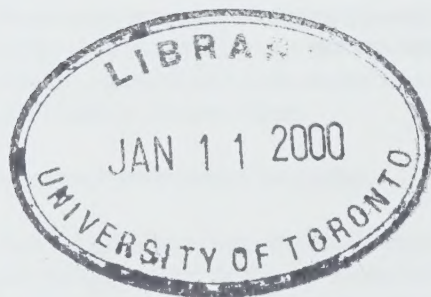
CS92-371-XPB

1. Sampling (Statistics).
 2. Canada -- Census, 1996 -- Methodology.
 3. Canada -- Census, 1996 -- Evaluation.
- I. Statistics Canada. II. Title. III. Échantillonnage et pondération. IV. Series.

HA741.5 1996 S25 1999 304.6'07'23
C99-988039-XE

How to Cite this Document

Statistics Canada. Sampling and Weighting.
1996 Census Technical Reports.
Ottawa: Minister of Industry, 1999.
Catalogue number 92-371-XPB.



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials.
ANSI Z39.48-1984



Table of Contents

	Page
List of Tables.....	ii
I. Introduction	1
II. Sampling in Canadian Censuses.....	3
A. The History of Sampling in the Canadian Census	3
B. The Sampling Scheme Used in the 1996 Census.....	4
C. Processing the Census Sample	5
III. Estimation from the Census Sample.....	7
A. Operational Considerations.....	7
B. Theoretical Considerations	7
C. Developing an Estimation Procedure for the Census Sample.....	8
D. The Two-step Generalized Least Squares Estimation Procedure (GLSEP)	9
IV. The Sampling and Weighting Evaluation Program.....	11
A. Sampling Bias Study	11
B. Evaluation of Weighting Procedures	11
C. Sample Estimate and Population Count Consistency Study.....	11
D. Sampling Variance	11
V. Sampling Bias.....	13
VI. Evaluation of Weighting Procedures.....	21
A. Weighting Area (WA) Formation	21
B. Evaluation of the Generalized Least Squares Estimation Procedure.....	22
1. Distribution of Weights	23
2. Discrepancy Between Population Counts and Sample Estimates	23
3. Discarding Constraints	25
4. Evaluation at Various Geographic Levels.....	27
VII. Sample Estimates and Population Count Consistency	29
A. Census Divisions (CDs)	29
B. Census Subdivisions (CSDs)	30
C. Census Tracts (CTs).....	30
D. Enumeration Areas (EAs)	31
E. Impact of the Changes to the Weighting Procedure in 1996	31
VIII. Sampling Variance	37
A. 1986 Census Sampling Variance Study	37
B. Sampling Variance and Bias With Generalized Least Squares Estimation.....	39
IX. Conclusion.....	41

	Page
Appendices	
A. Glossary of Terms	45
B. WA and EA Level Constraints Applied to the 1996 Census Weights.....	47
C. Additional Information on Statistics Used in Sampling Bias Study	49
D. 1986 Standard Error Adjustment Factors at National or Provincial Level and Percentiles of Weighting Area Level Factors.....	51
E. Products and Services	59
Bibliography	61
Regional Reference Centres	63

List of Tables

5.1 Summary Statistics at the Canada Level	16
5.2 Summary Statistics for Z Values at the Census Division (CD) Level.....	17
5.3 National and Regional Z Values.....	18
5.4 Comparison Between 1991 and 1996	19
6.1 Dwelling Count Distribution and Contiguity of 1996 Weighting Areas.....	21
6.2 Number of CSDs, CTs and FEDs that Respect WA Boundaries.....	22
6.3 Distribution of Household Weights.....	23
6.4 1996 Estimate/Population Discrepancies at the Canada Level	25
6.5 Frequency of Discarding WA Level Constraints in 1996	27
6.6 Percentage of the Characteristics With R Values Falling in Certain Ranges.....	28
7.1 Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CDs – 1996 and 1991 Censuses	32
7.2 Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CSDs – 1996 and 1991 Censuses	33
7.3 Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CTs – 1996 and 1991 Censuses.....	34
7.4 Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for EAs – 1996 and 1991 Censuses.....	35
8.1 Non-adjusted Estimates of Standard Errors of Sample Estimates.....	40

I. Introduction

The 1996 Census required the participation of the entire population of Canada, i.e. some 29 million people distributed over a territory of 9.2 million square kilometres. An endeavour of this magnitude represented a tremendous challenge. Although there are high quality standards governing the gathering and processing of the data, in spite of efforts aimed at reducing non-response, for example through the use of communications, it is not possible to eliminate all errors. An error does not necessarily imply a mistake as such, as some element of error is bound to result from decisions to control census costs.

Statistics Canada must explain the methods and concepts used to collect and process its data, and provide users with information about the quality of the data produced as well as other data characteristics which might limit their usefulness or interpretation. This report is aimed at informing users about the complexity of the data and any difficulties that could affect their use. It explains the theoretical framework and definitions used to gather the data, and describes unusual circumstances that could affect data quality. Moreover, the report touches upon data capture and edit and imputation, and deals with the chronological comparability of the data.

This report deals with sampling and weighting. It has been prepared by Peter Dick, Karen Switzer, Sylvain Thivierge and Patrick Mason, with the support of staff from two divisions in Statistics Canada: the Social Survey Methods Division and the Census Operations Division.

Users will find additional information on census concepts, variables and geography in the *1996 Census Dictionary* (Catalogue No. 92-351-XPE), and an overview of the complete census process in the *1996 Census Handbook* (Catalogue No. 92-352-XPE).

Sampling is an accepted practice in many aspects of life today. The quality of produce in a market may be judged visually by a sample before a purchase is made; we form opinions about people based on samples of their behaviour; we form impressions about countries or cities based on brief visits to them. These are all examples of sampling in the sense of drawing inferences about the “whole” from information for a “part”.

In a more scientific sense, sampling is used, for example, by accountants in auditing financial statements, in industry for controlling the quality of items coming off a production line, and by the takers of opinion polls and surveys in producing information about a population’s views or characteristics. In general, the motivation to use sampling stems from a desire either to reduce costs or to obtain results faster, or both. In some cases, measurement may destroy the product (e.g., testing the life of light bulbs) and sampling is therefore essential. The disadvantage of sampling is that the results based on a sample may not be as precise as those based on the whole population. However, when the loss in precision (which may be quite small when the sample is large) is tolerable in terms of the uses to which the results are to be put, the use of sampling may be cost effective. Furthermore, the reduction in the scale of a study achieved through using sampling may in fact lead to a reduction in errors from non-sampling sources, thus compensating to some extent for the loss of precision resulting from sampling.

The 1996 Census of Population made use of sampling in a variety of ways. It was used in ensuring that the quality of the Census Representative’s work in collecting questionnaires met certain standards; it was used in the control of the quality of coding responses during office processing; it was used in estimating both the amount of under-coverage and the amount of over-coverage which occurred for different reasons; it was used in evaluating the quality of census data. However, the primary use of sampling in the census was during the field enumeration when all but the basic census data were collected only from a sample of households. This report describes this last use of sampling and evaluates the effect of sampling on the quality of census data.

Chapter II reviews the history of the use of sampling in Canadian censuses and describes the sampling procedures used in the 1996 Census. Chapter III explains the procedures used for weighting up the sample data to the population level and provides operational and theoretical justifications for these procedures. In Chapter IV, the program of studies designed to evaluate the 1996 Census sampling and weighting procedures is presented, while Chapters V through VIII present the results of these studies. Chapter IX presents some conclusions on the weighting procedures used in 1996 and some suggestions for the 2001 Census.

II. Sampling in Canadian Censuses

In the context of a census of population, sampling refers to the process whereby certain characteristics are collected and processed only for a random sample of the dwellings and persons identified in the complete census enumeration. Tabulations that depend on characteristics collected only on a sample basis are then obtained for the whole population by scaling up the results for the sample to the full population level. Characteristics collected on all dwellings or persons in the census will be referred to as “basic characteristics” while those collected only on a sample basis will be known as “sample characteristics”.

A. The History of Sampling in the Canadian Census¹

Sampling was first used in the Canadian census in 1941. A housing schedule was completed for every tenth dwelling in each census subdistrict. The information from 27 questions on the separate Housing Schedule was integrated with the data in the personal and household section of the Population Schedule for the same dwelling, thus allowing cross-tabulation of sample and basic characteristics. Also in the 1941 Census, sampling was used at the processing stage to obtain early estimates of earnings of wage-earners, of the distribution of the working-age population and of the composition of families in Canada. In this case, a sample of every tenth enumeration area across Canada was selected and all population schedules in these areas were processed in advance.

Again in 1951, the Census of Housing was conducted on a sample basis. This time, every fifth dwelling (those whose identification numbers ended in a 2 or a 7) was selected to complete a housing document containing 24 questions. In the 1961 Census, persons 15 years of age and over in a 20% sample of private households were required to complete a population sample questionnaire containing questions on internal migration, fertility and income. Sampling was not used in the smaller censuses of 1956 and 1966.

The 1971 Census saw several major innovations in the method of census-taking. The primary change was from the traditional canvasser method of enumeration to the use of self-enumeration for the majority of the population. This change was prompted by the results of several studies in Canada and elsewhere (Fellegi, 1964; Hansen et al., 1959) that indicated that the enumerator had the effect of contributing significantly to the variance² of census figures in a canvasser census. Thus the use of self-enumeration was expected to reduce the variance of census figures through reducing the effect of the enumerator while, at the same time, giving the respondent more time and privacy in which to answer the census questions – factors which might also be expected to yield more accurate responses.

The second aspect of the 1971 Census that differentiated it from any earlier census was its content. The number of topics covered and the number of questions asked were greater than in any previous Canadian census. Considerations of cost, respondent burden and timeliness versus the level of data quality to be expected using self-enumeration and sampling led to a decision to collect all but certain basic characteristics on a one-third sample basis in the 1971 Census. In all but the more remote areas of Canada, every third private household received the “long questionnaire” which contained all the census questions, while the remaining private households received the “short questionnaire” containing only the basic questions covering name, relationship to head, sex, date of birth, marital status, mother tongue, type of dwelling, tenure, number of rooms, water supply, toilet facilities and certain coverage items. All

¹ More detailed information for specific censuses can be found in the Administrative Report, General Review, Summary Guide or Census Handbook of the appropriate census. References to these products can be found at the end of this report.

² The “variance” of an estimate is a measure of its precision. Variance is discussed more fully in Chapter VIII.

households in pre-identified remote enumeration areas and all collective dwellings³ received the long questionnaire. A more detailed description of the consideration of the use of sampling in the 1971 Census is given in "Sampling in the Census" (Dominion Bureau of Statistics, 1968).

The content of the 1976 Census was considerably less than that of the 1971 Census. Furthermore, the 1976 Census did not include the questions that cause the most difficulty in collection (e.g., income) or that are costly to code (e.g., occupation, industry and place of work). Therefore, the benefits of sampling in terms of cost savings and reduced respondent burden were less clear than for the 1971 Census. Nevertheless, after estimating the potential cost savings to be expected with various sampling fractions, and considering the public relations issues related to a reversion to 100% enumeration after a successful application of sampling in 1971, it was decided to use the same sampling procedure in 1976 as in 1971.

Most of the methodology used in the 1971 and 1976 Censuses was kept for the 1981 Census, except that the sampling rate was reduced from every third occupied private household to every fifth. Studies done at the time showed that the resulting reduction in data quality (measured in terms of variance) would be tolerable, and would not be significant enough to offset the benefits of reduced cost and response burden and would improve timeliness (see Royce, 1983). Twelve questions were asked on a 100% basis and an additional 34 questions were asked in the sample.

The 1986 Census was the first full mid-decade census. It was decided that only a full census could meet the growing need for local labour market data, a need made more pressing by the occurrence of a major recession (1981-82) since the previous census. However, in order to keep development costs as low as possible, a policy of minimum change was adopted. Unless there were compelling reasons not to do so, the 1981 Census questions and data collection and processing procedures were to be retained. Questions on eight subjects from the 1981 Census were not asked in 1986, while three new questions were added.

In 1991, the Census of Population included permanent and, for the first time, non-permanent residents – persons who hold student or employee authorizations, or Minister's permits, or who are refugee claimants. In order to identify non-permanent residents, a new question for the 1991 Census was designed, tested and added. In total, twelve new questions were added while four questions from 1986 were not asked in 1991. Two post-censal surveys were conducted following the collection on the 1991 Census. One of these surveys, the Health and Limitation Survey, collected information of the health and general well-being of Canadians. The second survey, the Aboriginal Peoples Survey, collected information on the Aboriginal population living on and off reserves. In addition, in 1991, there was a significant increase in the automation of data processing as well as in the way the products and services were delivered to the clients.

B. The Sampling Scheme Used in the 1996 Census

A wealth of information was collected from everyone in Canada on Census Day, 1996. Of note, in 1996, Census Day itself was moved from the traditional early June date to May 14, 1996. The bulk of the information was acquired on a sample basis. In all self-enumeration areas, a 1 in 5 sample of private occupied households was selected to receive a long questionnaire (Form 2B), containing all census questions. Basic questions on age, sex, marital status, mother tongue and relationship to the household reference person (Person 1) were asked on 100% basis while additional information on dwelling type, tenure and socio-economic questions were asked on a 20% basis.

³ A collective dwelling is a dwelling of a commercial, institutional or communal nature. Examples include hotels, hospitals, staff residences and work camps.

All dwellings in those areas enumerated by the canvasser method (generally remote areas or Indian reserves) received the Form 2B. All collective dwellings also received the Form 2B. However, the following persons in collective dwellings were not asked the sample questions:

- (a) inmates in correctional and penal institutions or jails;
- (b) patients in general hospitals, special care homes and institutions for the elderly, and chronically ill or psychiatric institutions;
- (c) children in orphanages and children's homes or young offenders' facilities.

Canadians stationed abroad (generally embassy or Armed Forces personnel) were given a Form 2C, which contained the same questions as the Form 2B except that housing questions were not included. However, questions about the person's usual place of residence in Canada were asked. Information on unoccupied private dwellings was recorded on a Form 2A.

The basic drop-off or delivery procedure required the Census Representative (CR) to pre-plan a route covering all dwellings in his or her enumeration area (EA) and then to visit each dwelling and leave a census questionnaire. The selection of the sample, i.e. the decision as to which type of questionnaire to leave at each occupied dwelling, was facilitated by the Visitation Record (VR), the document in which the CR listed each dwelling in his or her area. This document was printed so that every fifth line was shaded to signify that a Form 2B should be delivered. A random start was implemented by deleting either zero, one, two, three or four lines at the start of the VR according to whether the fifth, fourth, third, second or first dwelling in the EA was to be the first to receive the long questionnaire. Thereafter, the dwelling listed on each shaded line automatically received the long questionnaire. These procedures were spelled out in the CR's Manual and emphasized in his or her training in order to minimize the risk of any deviation from the specified procedure for selecting the sample.

In the 1996 Census, a major test on census operations was conducted in over 400,000 households in Eastern Ontario – including the Ottawa area – with a mail-out/mail-back questionnaire. Basically, all the households in the urban areas were listed on an address register and questionnaires were mailed to these dwellings with the sample households (2B) selected systematically from the Address Register. Respondents were expected to mail back the questionnaires; a small group of highly trained personnel followed up on the non-respondents.

In sampling terminology, the census sample design can be described as a stratified systematic sample of private occupied dwellings using a constant 1 in 5 sampling rate in all strata (EAs). As a sample of persons, it can be regarded as a stratified systematic cluster sample with dwellings as clusters. For a more detailed description of the concepts and terminology of sampling, see Sarndal, Swensson and Wretman, 1992.

C. Processing the Census Sample

Once the CR had obtained the completed questionnaire (Form 2A or 2B) from each dwelling in his or her area, and this work had been approved, the questionnaires were sent to one of seven regional processing sites for manual processing. Complete data for each EA were captured and stored on magnetic tapes. The questionnaires and magnetic tapes were then sent to Head Office Processing in Ottawa. Once there, checks were performed by computer for various inconsistencies in the data which required a manual review of the questionnaire to resolve. After all resulting updates to the data for an EA were completed, the data were reformatted and transferred to Edit and Imputation.

The data were loaded onto Edit and Imputation databases, organized by 2A (100%) and 2B (20%), with five regions for each database. The 2A databases contained the basic demographic characteristics for 100% of the population, while the 2B databases contained the data for the 20% sample questions. The data were processed through a series of customized modules, where all problems of invalid, inconsistent and missing data were resolved. The 2A databases were processed first, and a final 2A Canada Retrieval Database was created.

Once the 100% data were finalized, the data for the 20% sample questions were processed. Non-response 2B records were dropped from the 2B databases. A final 2B Canada Retrieval Database was created which contained both the 100% and 20% data for sampled households and persons only. The weights created using the 100% data (as described in Chapter III) were placed on this database.

III. Estimation from the Census Sample

Any sampling procedure requires an associated estimation procedure for scaling sample data up to the full population level. The choice of an estimation procedure is generally governed by both operational and theoretical constraints. From the operational viewpoint, the procedure must be feasible within the processing system of which it is a part; from the theoretical viewpoint, the procedure should minimize the sampling error of the estimates it produces. In the following two sections, the operational and theoretical considerations relevant to the choice of estimation procedures for the census sample are described.

A. Operational Considerations

Mathematically, an estimation procedure can be described by an algebraic formula that shows how the value of the estimator for the population is calculated as a function of the observed sample values. In small surveys that collect only one or two characteristics, or in cases where the estimation formula is very simple, it might be possible to calculate the sample estimates by applying the given formula to the sample data for each estimate required. However, in a survey or census in which a wide range of characteristics is collected, or in which the estimation formula is at all complex, the procedure of applying a formula separately for each estimate required is not feasible. In the case of a census, for example, every cell of every tabulation based on sample data at every geographic level represents a sample estimate which, under this approach, would require a separate application of the estimation formula. In addition, the calculation of each estimate separately would not necessarily lead to consistency between the various estimates made from the same census sample.

The approach taken in the census therefore (and in many sample surveys) is to split the estimation procedure into two stages: (a) the calculation of weights (known as the weighting procedure) and (b) the summing of weights to produce estimated population counts. Any mathematical complexity is then contained in step (a) which is performed just once, while step (b) is reduced to a simple process of summing weights which takes place at the time a tabulation is retrieved. Note that, since the weight attached to each sample unit is the same for whatever tabulation is being retrieved, consistency between different estimates based on sample data is assured.

B. Theoretical Considerations

For a given sample design and a given estimation procedure, one can, from sampling theory, make a statement about the chances that a certain interval will contain the unknown population value being estimated. The primary criterion in the choice of an estimation procedure is the minimization of the width of such intervals so that these statements about the unknown population values are as precise as possible. The usual measure of precision for comparing estimation procedures is known as the standard error. Provided that certain relatively mild conditions are met, intervals of plus or minus two standard errors from the estimate will contain the population value for approximately 95% of all possible samples.

As well as minimizing the standard error, a second objective in the choice of the estimation procedure for the census sample is to ensure, as far as possible, that sample estimates for basic (i.e. 2A) characteristics are consistent with the corresponding known population values. Fortunately, these two objectives are usually complementary in the sense that the sampling error tends to be reduced by ensuring that the sample estimates for certain basic characteristics are consistent with the corresponding population figures. While this is true in general, however, forcing sample estimates for basic characteristics to be consistent with corresponding population figures for very small subgroups can have a detrimental effect on the standard error of estimates for the sample characteristics themselves.

In the absence of any information about the population being sampled other than that collected for sample units, the estimation procedure would be restricted to weighting the sample units inversely to their probabilities of selection (e.g., if all units had a 1 in 5 chance of selection, then all selected units would receive a weight of 5). In practice, however, one almost always has some supplementary knowledge about the population (e.g., its total size and, possibly, its breakdown by a certain variable – perhaps by province). Such information can be used to improve the estimation formula so as to produce estimates with a greater chance of lying close to the unknown population value. In the case of the census sample, a large amount of very detailed information about the population being sampled is available in the form of the basic 100% data at every geographic level. We can take advantage of this wealth of population information to improve the estimates made from the census sample. However, this information can also be an embarrassment in the sense that **it is impossible to make the sample estimates for basic characteristics consistent with all the population information at every geographic level.** Differences between sample estimates and population values become visible when a cross-tabulation of a sample variable and a basic variable is produced. The tabulation has to be based on sample data, with the result that the marginal totals for the basic variable are sample estimates that can be compared with the corresponding population figures appearing in a different tabulation based on 100% data. They will not necessarily agree exactly.

C. Developing an Estimation Procedure for the Census Sample

Given that a weight has to be assigned to each unit (person, family or household) in the sample, the simplest procedure would be to give each unit a weight of 5 (because a 1 in 5 sample was selected). Such a procedure would be simple and unbiased⁴ and, if nothing but the sample data were known, it might be the optimum procedure. However, although we know that the sample will contain almost exactly one fifth of all households (excluding collective households and those in canvasser areas), one cannot be certain that it will contain exactly one fifth of all persons, or one fifth of each type of household, or one fifth of all females aged 25-34, and so on. Therefore, this procedure would not ensure consistency even for the most important subgroups of the population. For large subgroups, these fractions should be very close to one-fifth, but for smaller subgroups they could differ markedly from one-fifth. The next most simple procedure would be to define certain important subgroups (e.g., age-sex groups within the province) and, for each subgroup, to count the number of units in the population in the subgroup (N) and the number in the sample (n) and to assign to each sample unit in the subgroup a weight equal to N/n.

For example, if there were 5,000 males aged 20-24 enumerated in Prince Edward Island, and 1,020 of these fell in the sample households, then a weight of $5,000/1,020 = 4.90$ would be assigned to each male aged 20-24 in the sample in Prince Edward Island. This would ensure that whenever sex and age in five-year groups were cross-classified against a sample characteristic for Prince Edward Island, the marginal total for the male 20-24 age-sex group would agree with the population total of 5,000. This type of estimation procedure is known as “ratio estimation”. It should be noted in this particular example that a weight of 5 would result in a sample estimate of 5,100 ($1,020 \times 5$). The estimation procedure used prior to the 1991 Census was a generalization of ratio estimation called the raking ratio estimation procedure (RREP). For more details on the RREP, see the *User's Guide to the Quality of 1986 Census Data: Sampling and Weighting* (Statistics Canada, 1990) as well as Brackstone and Rao, 1979.

It was decided to use an alternative estimation procedure called the two-step Generalized Least Squares Estimation Procedure (GLSEP) for the 1991 Census. This was done to achieve a higher level of agreement between population

⁴ “Unbiased” means that the average of the estimates obtained by this procedure, over all possible samples, would equal the true population value.

counts and the corresponding estimates at the EA level than was possible with the RREP. The standard errors of the estimates under GLSEP for small geographic areas were also reduced. In addition, the GLSEP allowed a single weight to be determined for each sampled household which was used to produce estimates for both person and household characteristics. By contrast, with the RREP, it was necessary to use different weights to produce estimates for household and person characteristics: this contributed to inconsistencies.

With the GLSEP, the initial weights of approximately 5 were adjusted as little as possible for individual households such that there was perfect agreement between the estimates and the population counts for as many of the basic characteristics as possible that are listed in Appendix B (these will be called constraints). It was required that this perfect agreement be achieved at the weighting area (WA) level. Each WA contained on average seven sampled EAs. More information on WAs is given in Chapter VI, Section A of this report. It can be shown that the GLSEP is a regression estimator. For more details on regression estimators, a good reference is Sarndal, Swensson and Wretman, 1992.

D. The Two-step Generalized Least Squares Estimation Procedure (GLSEP)

The weight calculations are carried out independently in each weighting area (WA). Some of the constraints (both at the EA and WA levels) listed in Appendix B have to be discarded for each WA and hence population/estimate agreement cannot be guaranteed for all constraints. Constraints are initially discarded at the WA level because:

- they apply to less than 20 households (these will be called **small** constraints);
- they are redundant (these will be called **linearly dependent** (LD) constraints);
- they are nearly redundant (these will be called **nearly linearly dependent** (NLD) constraints).

A redundant constraint would be, for example, the total number of females, since constraints are already present when using total number of persons and total number of males. A **linearly dependent constraint** occurs when any two (or more) constraints guarantee that the third will be automatically satisfied. An example of a **nearly redundant constraint** can be seen by considering the constraints **marital status = single** and **household size = 1**. If most, but not all, persons in households of size 1 are single, then the two constraints are almost equal and one constraint can be considered NLD. The LD constraints were discarded to increase the computational efficiency of the weighting algorithm. The small and NLD constraints were discarded because, otherwise, the estimates might become unstable and have large standard errors.

After small, LD and NLD constraints are discarded at the WA level, the calculation of the GLSEP weights takes place in two steps. In the first step, the initial weights, which equal the reciprocal of the EA household sampling fraction, are adjusted individually for each EA. Note that some constraints, when they were not discarded at the WA level, may be discarded because of smallness or linear dependence at the EA level. The remaining constraints that have not been discarded in the EA are sorted by the number of households that they apply to at the EA level. The constraints are then split into two groups, with the even-numbered constraints in one and the odd-numbered constraints in the other. The GLSEP weights are calculated at the EA level for each group of constraints. Sometimes, the estimation procedure will produce very small weights (less than zero) or very large weights (greater than 25) in order to obtain the necessary agreement for certain constraints. These weights, which are called “outlier” weights, are considered undesirable. Consequently, when this occurs, the constraints causing them are identified and discarded, and the weights are recalculated. Finally, the weights for the two groups of constraints are averaged together for each sampled household to produce the first step weights for each EA.

The weights produced in the first step are used as initial weights in the second step, where they are adjusted so that agreement is obtained between sample estimates and population counts at the WA level. All constraints not identified

as small, LD or NLD at the WA level are used. If any outlier weights are produced, the constraints causing them are identified and discarded, and the final weights are recalculated. Although the second step distorts somewhat the agreement obtained for estimates at the EA level in the first step, the final EA level estimates are still closer to the population counts than they would have been had the first step not been done. For a more detailed explanation of the calculation of the weights, see Bankier, Rathwell and Majkowski, 1992.

In 1996, two separate estimation runs were completed. The second estimation run was deemed necessary when some unacceptable discrepancies were detected for the variable “common-law status” and for “household size” from the first estimation run. The second run included a new constraint “common-law status (equals yes)” and an adjustment to the initial weight. For the first run, the initial household weight equalled the inverse of the household level sampling fraction. However, in the second run, these initial weights were adjusted so that the estimated number for each household size agreed with the population count at the WA level. Further information can be found in Bankier, Houle and Luc, 1997.

GLSEP weights were calculated only for private households that received the long census questionnaire in a sampled EA (1/5 of the private households were sampled, the other 4/5 were not). Private households that received a short questionnaire in a sampled EA received a weight of 0. All private households in non-sampled EAs received a weight of 1, because 100% of the respondents in such areas provide information for the Form 2B. Collective households also received a weight of 1.

IV. The Sampling and Weighting Evaluation Program

The Sampling and Weighting Evaluation Program was designed to determine the effect of sampling and weighting on the quality of census sample data. To this end, four studies were carried out to measure the quality of the census sample data and estimates and to provide information relevant to the planning of future censuses. These studies were:

- (a) an examination of sampling bias;
- (b) an evaluation of the weighting procedures;
- (c) an evaluation of sample estimate and population count consistency;
- (d) a study to evaluate the sampling variance for various 20% sample characteristics.

In the remainder of this chapter, these four studies are briefly described. Chapters V through VIII present the results of these studies.

A. Sampling Bias Study

Bias can be introduced into responses to any survey from a number of sources. The objective of this study was to determine if responses to basic questions on Forms 2B were biased in any way and to identify, if possible, the causes of any observed bias.

B. Evaluation of Weighting Procedures

The objective of this study was to evaluate the performance of the GLSEP. The level of agreement between the sample estimates and population counts for the constraints over all WAs in Canada was examined. The number and type of constraints discarded at the WA level as well as the reason for them being discarded were studied to explain observed inconsistencies. In addition, the distribution of the GLSEP weights as well as differences between 1991 results and 1996 results were studied.

C. Sample Estimate and Population Count Consistency Study

This study examined the level of agreement (consistency) between sample estimates and population counts for a wide variety of basic characteristics. This consistency was studied for various geographic areas other than WAs. Comparisons are also made between the consistency achieved in 1996 and 1991 for these characteristics.

D. Sampling Variance

The “variance” of an estimate is a measure of its precision. Estimates of variance for estimators using simple weights of 5 and assuming simple random sampling are relatively inexpensive to calculate. However, estimates of variance for census estimators taking into account the sample design and estimation techniques used are very expensive to calculate. **Adjustment factors** were calculated for the 1986 Census; they are the ratios of the estimates of the

standard errors (the square roots of the variances) for census estimates to the simple estimates of the standard errors. An estimate of the standard error of a census estimate for any characteristic in any geographic area can then be obtained by multiplying the simple estimate of the standard error by the appropriate adjustment factor. In addition, the study discussed how these estimates of the standard error may not be accurate because of the bias introduced into the process by the sample, the data processing and the estimation procedure.

V. Sampling Bias

Estimates based on a sample survey are subject to sampling errors. One type of sampling error arises from the variability in the population. This variability manifests itself in different samples producing different estimates, none of which will necessarily equal the true population value. The estimates will equal the true population value on average **provided that there is no bias in the sample**. With the presence of a bias, the true population value tends to be overestimated or underestimated. Unfortunately, a bias is often difficult to eliminate completely. In the Census of Population, a bias can be introduced into the responses from a variety of sources. These include coverage errors, non-response bias, response bias (e.g., respondents answering differently on the Form 2B than on the 2A), Census Representative (CR) errors (e.g., not selecting the sample according to specifications), processing errors, and so on.

The Census of Population gives a unique opportunity to examine closely the bias of the basic characteristics. Essentially, there are two estimates of the same basic characteristic: the complete population count and the estimate from the 1 in 5 sample estimate. The purpose of the Sampling Bias Study was to examine the bias in the responses to the basic questions on Forms 2B. The sample estimates were produced by multiplying the sample counts at the EA level by simple weights – equal to the inverse of the EA household sampling fraction (approximately 5) – and then summing to the appropriate geographic level⁵.

Initially, the results were summed to the national level. In Table 5.1, the national level results for the population count, the estimate, the difference between the estimate and the population count, the discrepancy and the standard error of the estimate for 32 basic characteristics are displayed. In addition, the Z value – which provides a test of statistical significance – is also shown: these are discussed in more detail below. For 22 characteristics, the Z value is outside the expected range of -2 to 2. For instance, the Z values for males, ages 20-29, ages 75 and over and males over age 15 are all smaller than -2. This indicates that the sample, compared to the population, is under-represented for these groups at the national level. However, the Z values for females, total population, ages 0-14 and married, are greater than 2, indicating that these characteristics are over-represented in the sample. Also, one-person and six-or-more-person households were under-represented while households of size 2 to 5 were over-represented. Generally the sample contains too many children, middle-aged persons, females and married people while it does not contain enough young adults, elderly persons, males and unmarried persons.

Those differences which were statistically significant at the 5% level in Table 5.1 were identified using the Z value

$$Z^{(0)} = \frac{\hat{X}^{(0)} - X}{\sqrt{V(\hat{X}^{(0)})}} \quad (1)$$

which is discussed in more detail in Appendix C. The statistic was computed using data with imputations for each of the 281 census divisions (CDs) in Canada and for the set of 32 characteristics where $\hat{X}^{(0)}$ is an estimate based on the 2B sample using simple weights while the known 2A population count X and $V(\hat{X}^{(0)})$ is the sampling variance of the estimator $\hat{X}^{(0)}$. The $Z^{(0)}$ values, for the 281 CDs, should approximately follow a normal distribution with mean 0 and variance 1 if a simple random sample – or a systematic sample – of households was selected unbiasedly from each EA and was not affected by processing.

⁵ These simple estimates were used instead of the GLSEP estimates because the GLSEP tends to mask the sampling bias by forcing estimates of basic characteristics to equal population counts.

Table 5.1 presents the results at the national level, but this does not explain whether a few negatively-biased results are offsetting other positively-biased results. To examine this question, a summary of the bias statistics in the 281 CDs are shown in Table 5.2. This table shows the mean, the standard error and the test result (of assessing if the true mean of the Z values is zero) for the Z values. This hypothesis was rejected if $|T| > 1.96$ of the Z values for the basic 32 characteristics. Generally, the same characteristics that were detected as being biased in Table 5.1 – at the national level – are also found to be biased when the entire set of CDs is examined. However, certain changes are also notable. For instance, males at the national level were seen to be under-represented in the sample, but when the CDs are examined this bias is not detected. Since Table 5.2 **assumes all CDs are the same size**, we can conclude that larger (urban) CDs have a negatively-biased estimate of males while smaller (rural) CDs probably have an over-representation of males. Conversely, for households of size 5, a bias was detected nationally, but when the sampled CDs are examined, we can see that they are shown to be over-represented in the sample. The explanation is that, while there are very few CDs with either large over- or under-representations, there is a small, but persistent, over-representation throughout the sample.

Table 5.2 also shows the percentage of CDs with Z values less than -2, the percentage greater than 2 and the percentage greater than 0. If the sample was unbiased, we would expect that approximately 2.5% of the CDs would be either less than -2 or greater than 2, while we would expect 50% of the CDs to have Z values larger than 0. Clearly some characteristics demonstrate that the bias is widespread through almost all CDs. For instance, married persons are over-represented in 82% of the CDs. Furthermore, 13% of the CDs have Z values for married persons of over 2, while only 0.4% of the CDs – meaning 1 CD out of 281 – has a Z value less than -2. The census sample clearly contains too many married persons. On the other hand, only 26% of the CDs have an over-representation – a Z value greater than 0 – of one-person households and, in addition, not even one CD had a Z value greater than 2 for one-person households.

As mentioned in the Introduction, there are many possible explanations for the observed differences between the sample estimates based on simple weights and the population counts. One possible cause is the 0.8% of the total households (i.e. 86,183) which were missed/refusal households in the 1996 Census. This compares with 2.6% of the total (i.e. 253,156) in the 1991 Census. These were households which either completely refused to answer the questions or for which the Census Representative (CR) was unable to get any information, because the members of the household were absent during the census-taking period or had moved on or after Census Day without responding. Sometimes the CR was able to determine the number of persons, but usually all other responses had to be imputed for these households. Of the missed/refusal households in 1996, 32,820 (38% of 86,183) were sampled households, while we would have expected that about 20% would be from the sampled households if non-responses were equally likely for 2A and 2B households. In addition, 3,358 of the sampled households, while not missed/refusal (i.e. they provided some responses to the basic questions), provided no answers to the questions asked on a sample basis. During data processing, these $32,820 + 3,358 = 42,179$ sampled households with complete non-response to the sampled questions were removed from the sample (i.e. they were converted from Form 2B to Form 2A households so that they became non-sampled households) and the responses to the basic questions only were imputed. This procedure of converting sampled households to non-sampled households is known as 2A/2B document conversion. It is possible that the missed/refusal households and the households without responses to the sample questions had different characteristics (e.g., they could have been smaller) for known responses than other households. Thus converting Forms 2B to Forms 2A could bias the sample. Also, if the imputation system had a tendency to impute certain characteristics for missed/refusal households more often than for other types of households, this would have caused sample estimate and population count discrepancies as well, since only non-sampled households would be affected.

The geographic variation of the bias was also studied. Table 5.3 shows the resulting Z values for 32 characteristics for Canada, the East region, Quebec, Ontario and the West region. While it is difficult to draw definitive conclusions from this table – since the sample sizes vary to a large extent – some patterns become evident. Quebec has a strikingly different pattern for males – the Quebec sample has males over-represented, while the rest of the country has males

under-represented in the sample. However, Ontario is different from the other regions in terms of total population. Quebec also has an over-representation in the sample of single persons and common-law status, while the rest of the country has an under-representation; but this might be more a reflection of the difference in the recording of marital status in Quebec. Notice that the age groups largely agree across the regions, in terms of the direction of the bias except for some minor differences for some middle-aged persons. The sample for persons over age 15 is fairly consistent in all regions, with males being under-represented and females over-represented (except in Ontario, where females are very slightly under-represented). Finally, the household sizes are also fairly consistent in the pattern of bias. Overall, while there are some differences in the regional bias, there are also a lot of patterns that are shared.

Finally, Table 5.4 briefly highlights some of the results from comparing the 1991 Sampling Bias Study with the 1996 study. When comparing the results at the CD level, it can be seen that only for males and single persons did the direction of the bias change. In 1991, these characteristics were over-represented in the sample, while in 1996 they were under-represented. When the number of CDs with Z values greater than 0 are compared, the difference in the males between the two samples is quite striking. In 1991, 63% of the CDs had males over-represented in the sample, while in 1996, 47% were over-represented. The other major change is in the number of CDs under-representing single persons: in 1991, 56% of the CDs over-represented single persons compared to 47% in 1996. However, in terms of ages, the two samples seem to be remarkably similar in pattern – only persons aged 35-44 were over-represented at a higher percentage of CDs in 1991 than in the 1996 sample.

In summary, there are biases present in the 2B sample and they are similar to the biases found in the 1991 Census sample. It is not clear where the bias comes from and it is not clear what can really be done to reduce it. However, the chosen weighting system – GLSEP – will largely correct the biases for those characteristics selected as constraints. It is an open question as to the resulting bias on the characteristics that are not retained as constraints.

Table 5.1 Summary Statistics at the Canada Level

Characteristic	Count	Estimate ¹	Difference ²	Disc. ³	S.E. ⁴	Z value ⁵
Males	13,717,654	13,694,786	(22,868)	-0.17	5,752	-3.98
Females	14,176,680	14,222,665	45,985	0.32	5,552	8.28
Total population	27,894,334	27,917,451	23,117	0.08	8,227	2.81
Age 0-4	1,858,332	1,874,111	15,779	0.85	3,073	5.14
Age 5-9	1,932,023	1,950,728	18,705	0.97	3,120	6.00
Age 10-14	1,939,776	1,957,694	17,918	0.92	3,125	5.73
Age 15-19	1,903,023	1,907,732	4,709	0.25	3,074	1.53
Age 20-24	1,840,654	1,816,301	(24,353)	-1.32	3,013	-8.08
Age 25-29	1,971,123	1,953,292	(17,831)	-0.90	3,053	-5.84
Age 30-34	2,405,559	2,401,580	(3,979)	-0.17	3,317	-1.20
Age 35-39	2,486,060	2,482,136	(3,924)	-0.16	3,339	-1.18
Age 40-44	2,268,423	2,273,674	5,251	0.23	3,177	1.65
Age 45-49	2,050,229	2,059,233	9,004	0.44	3,040	2.96
Age 50-54	1,581,484	1,589,751	8,267	0.52	2,707	3.05
Age 55-59	1,271,221	1,269,086	(2,135)	-0.17	2,448	-0.87
Age 60-64	1,157,926	1,160,459	2,533	0.22	2,338	1.08
Age 65-74	1,991,721	1,996,303	4,582	0.23	3,068	1.49
Age 75 and over	1,236,780	1,225,372	-11,408	-0.92	2,332	-4.89
Single persons	12,779,218	12,741,878	(37,340)	-0.29	7,320	-5.10
Married persons	11,537,475	11,628,813	91,338	0.79	6,076	15.03
Widowed persons	1,303,304	1,291,501	(11,803)	-0.91	2,130	-5.54
Divorced persons	1,605,136	1,591,530	-13,606	-0.85	2,612	-5.21
Separated persons	669,201	663,729	-5,472	-0.82	1,675	-3.27
Common-law = yes	1,770,338	1,768,774	-1,564	-0.09	3,568	-0.44
Males ≥ 15	10,781,073	10,732,804	-48,269	-0.45	4,449	-10.85
Females ≥ 15	11,383,130	11,402,113	18,983	0.17	4,006	4.74
One-person households	2,584,348	2,558,041	-26,307	-1.02	2,524	-10.42
Two-person households	3,385,597	3,397,657	12,060	0.36	3,011	4.00
Three-person households	1,804,304	1,809,076	4,772	0.26	2,435	1.96
Four-person households	1,813,493	1,825,159	11,666	0.64	2,378	4.91
Five-person households	737,751	740,921	3,170	0.43	1,640	1.93
Six-or-more-person households	334,207	327,786	-6,421	-1.92	1,124	-5.71

¹ Based on simple weights

² Difference: estimate-count

³ Disc.: discrepancy (100*[difference-count]/count)

⁴ S.E.: standard error of the simple weight estimate

⁵ Z value: (estimate-bias-count)/S.E.

Table 5.2 Summary Statistics for Z Values at the Census Division (CD) Level

Characteristic	Mean	Std ¹	T ²	% CDs (Z > 2)	% CDs (Z < -2)	% CDs (Z > 0)
Males	-0.07	1.08	-1.10	2.1	3.6	47.3
Females	0.55	1.08	8.50	8.2	0.4	69.7
Total population	0.32	1.06	4.98	6.0	0.7	59.8
Age 0-4	0.33	1.06	5.26	6.4	2.5	60.1
Age 5-9	0.33	1.04	5.29	4.3	0.7	61.2
Age 10-14	0.29	0.99	4.87	3.9	0.7	60.1
Age 15-19	0.10	0.98	1.73	3.6	0.7	54.8
Age 20-24	-0.38	1.04	-6.20	1.1	7.1	37.4
Age 25-29	-0.13	1.07	-2.03	1.8	2.8	48.4
Age 30-34	0.15	1.02	2.38	3.2	2.8	60.1
Age 35-39	0.03	0.98	0.60	2.1	1.8	53.7
Age 40-44	0.13	1.01	2.10	3.6	0.7	53.4
Age 45-49	0.22	1.10	3.29	5.3	2.1	59.1
Age 50-54	0.13	1.03	2.12	4.3	2.1	55.5
Age 55-59	-0.05	0.99	-0.78	1.8	2.1	49.8
Age 60-64	0.05	1.10	0.82	2.5	4.6	53.4
Age 65-74	-0.02	1.03	-0.40	1.8	2.1	50.5
Age 75 and over	-0.43	1.01	-7.11	0.4	7.8	31.0
Single persons	-0.13	1.10	-1.91	1.1	3.9	46.6
Married persons	0.87	1.03	14.15	12.5	0.4	81.9
Widowed persons	-0.41	0.96	-7.06	0.4	4.3	31.7
Divorced persons	-0.23	1.02	-3.78	0.7	6.0	42.7
Separated persons	-0.21	0.94	-3.68	1.4	2.1	39.9
Common-law = yes	-0.02	0.91	-0.33	0.7	1.4	52.0
Males ≥ 15	-0.39	1.19	-5.50	1.8	6.8	36.7
Females ≥ 15	0.34	1.08	5.22	6.4	1.1	62.3
One-person households	-0.69	1.02	-11.37	0.0	9.6	26.3
Two-person households	0.20	1.04	3.18	5.3	2.1	55.9
Three-person households	0.15	0.98	2.54	3.2	0.7	54.8
Four-person households	0.30	1.12	4.44	3.2	2.5	62.3
Five-person households	0.16	0.96	2.74	1.1	1.4	56.9
Six-or-more-person households	-0.26	1.03	-4.18	1.4	5.3	40.9

¹ Std: standard deviation of the Z values² T : Student statistic for testing that mean = 0

Table 5.3 National and Regional Z Values

Characteristic	Canada	East	Quebec	Ontario	West
Males	-3.98	-0.37	2.48	-5.69	-2.91
Females	8.28	3.87	7.57	1.82	4.29
Total population	2.81	2.34	6.94	-2.75	0.89
Age 0-4	5.14	0.33	4.37	2.10	2.91
Age 5-9	6.00	3.75	4.67	1.42	3.26
Age 10-14	5.73	1.61	2.88	2.88	3.81
Age 15-19	1.53	1.03	2.03	0.24	0.15
Age 20-24	-8.08	-2.46	-2.35	-4.78	-5.97
Age 25-29	-5.84	-0.81	-0.26	-6.31	-2.84
Age 30-34	-1.20	2.15	1.78	-3.51	-0.92
Age 35-39	-1.18	0.36	-0.52	-0.98	-0.77
Age 40-44	1.65	1.30	2.10	-0.79	1.30
Age 45-49	2.96	1.47	2.15	1.22	1.31
Age 50-54	3.05	-0.52	2.29	2.19	1.29
Age 55-59	-0.87	1.41	-0.33	-1.96	0.20
Age 60-64	1.08	0.11	3.02	-1.00	0.26
Age 65-74	1.49	-0.68	-1.02	2.93	0.70
Age 75 and over	-4.89	-3.54	-2.73	-1.46	-2.95
Single persons	-5.10	-0.58	0.87	-6.55	-2.69
Married persons	-15.03	6.10	9.16	7.12	8.03
Widowed persons	-5.54	-2.66	-2.01	-2.98	-3.61
Divorced persons	-5.21	-1.98	-0.98	-3.60	-3.85
Separated persons	-3.27	-1.51	-0.30	-2.28	-2.26
Common-law = yes	-0.44	-0.54	1.09	-1.23	-0.74
Males ≥ 15	-10.85	-2.31	-1.27	-9.23	-7.19
Females ≥ 15	4.74	3.04	6.27	-0.32	1.75
One-person households	-10.42	-5.02	-8.38	-3.20	-5.01
Two-person households	4.00	1.70	0.68	3.32	2.17
Three-person households	1.96	0.15	2.79	0.55	0.26
Four-person households	4.91	2.76	4.32	1.17	2.29
Five-person households	1.93	-0.02	2.43	-0.21	1.71
Six-or-more-person households	-5.71	-0.74	-0.80	-5.18	-3.12

Table 5.4 Comparison Between 1991 and 1996

Characteristic	1996			1991		
	T-Test	Mean	% CD > 0	T-Test	Mean	% CD > 0
Males	-1.1	-0.1	47	5.1	0.3	63
Females	8.5	0.6	70	12.0	0.8	78
Total population	5.0	0.3	60	11.2	0.7	76
Age 0-4	5.3	0.3	60	5.2	0.3	63
Age 5-9	5.3	0.3	61	7.6	0.5	68
Age 10-14	4.9	0.3	60	8.2	0.5	64
Age 15-19	1.7	0.1	55	3.4	0.2	59
Age 20-24	-6.2	-0.4	37	-3.6	-0.2	43
Age 25-29	-2.0	-0.1	48	-2.5	-0.2	46
Age 30-34	2.4	0.2	60	3.2	0.2	60
Age 35-39	0.6	0.0	54	3.0	0.2	62
Age 40-44	2.1	0.1	53	4.3	0.3	61
Age 45-49	3.3	0.2	59	4.3	0.3	62
Age 50-54	2.1	0.1	56	2.3	0.1	58
Age 55-59	-0.8	-0.1	50	-0.1	0.0	51
Age 60-64	0.8	0.1	53	1.0	0.1	51
Age 65-74	-0.4	0.0	51	-0.2	-0.0	49
Age 75 and over	-7.1	-0.4	31	-6.6	-0.4	35
Single persons	-1.9	-0.1	47	3.2	0.2	56
Married persons	14.2	0.9	82	17.4	1.1	85
Widowed persons	-7.1	-0.4	32	-6.1	-0.4	37
Divorced persons	-3.8	-0.2	43	-5.2	-0.3	38
Separated persons	-3.7	-0.2	40	-5.2	-0.3	39
One-person households	-11.4	-0.7	26	-17.1	-1.0	15
Two-person households	3.2	0.2	56	3.1	0.2	58
Three-person households	2.5	0.2	55	3.6	0.2	59
Four-person households	4.4	0.3	62	7.6	0.5	69
Five-person households	2.7	0.2	57	4.6	0.3	58
Six-or-more-person households	-4.2	-0.3	41	-1.4	-0.1	43

VI. Evaluation of Weighting Procedures

A. Weighting Area (WA) Formation

The first stage of the weighting procedures was the formation of WAs. A WA is the smallest geographic area for which agreement for characteristics of the population between certain sample and population counts can be ensured. Larger WAs allow population/estimate consistency to be achieved for more population characteristics but at the expense of small area population/estimate consistency. Therefore, when considering the WA size (number of households), we must acknowledge the trade-off between the need for agreement between the sample estimate and the population counts for small areas, and the need for this agreement for a wide variety of characteristics. For the 1996 Census, a WA was formed by grouping together enumeration areas (EAs) to adhere to the following conditions:

- (a) A WA must respect the boundaries of census divisions (CDs).
- (b) A WA should contain a population of between 800 and 3,000 households.
- (c) A WA should, where possible, respect certain high-level boundaries, giving priority in order of importance to census subdivisions (CSDs), then to census tracts (CTs) and last to federal electoral districts (FEDs).
- (d) A WA should be made up of contiguous EAs (i.e. be connected) and be as compact as possible.

Since the Generalized Least Squares Estimation Procedure (GLSEP) is performed independently within WAs, agreement between sample estimates and population counts is ensured only for those geographic areas which contain only entire WAs. Agreement is not ensured for geographic areas which are completely contained within a part of one WA or which contain parts of different WAs.

The following table shows the size distribution (in households) of the WAs.

Table 6.1 Dwelling Count Distribution and Contiguity of 1996 Weighting Areas

Dwellings	WA
0- 999	4
1,000-1,499	1,686
1,500-1,999	2,213
2,000-2,499	1,417
2,500-2,999	560
3,000+	61
Total	5,941

We observe from the above table that 5,876 (98.9%) of the 1996 WAs are within the desired range of 1,000 to 3,000 households. Also, 5,888 (99.1%) of the WAs were contiguous. Most of the remaining WAs were non-contiguous as a result of either an EA or CSD being non-contiguous.

The following table looks at geographic areas in terms of the number of geographic boundaries that are respected by WAs.

Table 6.2 Number of CSDs, CTs and FEDs that Respect WA Boundaries

Description	1996 Geographic Boundaries		
	CSD	CT	FED
Geographic areas containing only one whole WA or part of only one WA	5,373 (89.8%)	4,396 (73.5%)	0 (0%)
Geographic areas containing more than one whole WA	329 (5.5%)	313 (5.2%)	46 (15.6%)
Geographic areas that cross at least one WA boundary	282 (4.7%)	1,272 (21.3%)	249 (84.4%)
Total	5,984	5,980	295

The table indicates that only 282 (4.7%) of the CSDs do not respect WAs, 1,272 (21.3%) of CTs do not respect WAs and 249 (84.4%) of FEDs do not respect WAs.

For more information about weighting areas and their delineation, see Kruszynski, 1999.

B. Evaluation of the Generalized Least Squares Estimation Procedure

The 1991 weighting system was used in 1996 with very few changes to the software. While this implied little overall change to the system, the parameters – such as the defined constraints – could be modified. The aim was to retain a higher proportion of the constraints than were kept in 1991. In addition, due to budgetary reasons, a number of the 100% characteristics used in 1991 were eliminated.

During the first production run, constraints related to household sizes were often dropped at the WA level, thus permitting a population/estimate difference to be generated for these constraints. In addition, there was an important discrepancy of 2.6% for **common-law = yes** (which was not used as a constraint in the first run). These observations suggested that the weighting system was not completely successful in eliminating biases in the census sample. The implication was that the regression weights could then yield biased estimates for 2B characteristics. For these reasons, it was decided to run the weighting system a second time as described in Chapter III, Section D of this report.

To reduce the discrepancy for **common-law = yes**, this characteristic was added as a constraint in the second run. In addition, to help reduce the household size discrepancies, the initial weights were adjusted to agree with the population counts for the various household sizes at the WA level. In the first run, the initial weights for all the households in the same EA was simply the number of households in the EA divided by the number of sampled households in the EA: these weights are the **initial simple weights**. In the second run, the initial simple weights were adjusted to agree with the estimated number of households – for each of the six (1, 2, 3, 4, 5, 6+) household sizes – at the WA level. Thus, initially, prior to running the regression weighting system, there was agreement at the WA level between the sample estimates and the population counts for the six household sizes. These adjusted weights are called **initial poststratified weights**. The regression weighting system was then applied to these weights. Further details can be found in Bankier, Houle and Luc, 1997.

The evaluation of the weighting system is divided into four parts. Firstly, the distribution of the final weights is examined. Secondly, the discrepancies between the population counts and the final regression estimates are displayed and compared to comparable figures from 1991. Thirdly, the constraints that were discarded in the initial run are briefly described, and then complete details on the constraints discarded in the second run are displayed. Finally, the performance of the weighting system is evaluated at the various geographic levels.

1. Distribution of Weights

Table 6.3 shows the distribution of the weights for the two 1996 production runs and the corresponding 1991 distribution. Note that in order to retain more constraints than in 1991, it was decided to permit weights in the range (0.1) which implied that weights in 1996 are in the range of (0.25) instead of the 1991 range of (1.25).

Table 6.3 Distribution of Household Weights

Weight Range	1991		1996 – Second		1996 – First	
	Number	%	Number	%	Number	%
< 1.00	0	-	13,833	0.7	13,611	0.6
1.00 - 1.99	58,934	3.1	70,195	3.4	68,116	3.3
2.00 - 2.99	170,352	8.9	204,749	9.8	198,052	9.5
3.00 - 3.99	338,334	17.7	383,921	18.4	378,736	18.1
4.00 - 4.99	432,603	22.6	461,281	22.1	465,983	22.3
5.00 - 5.99	369,274	19.3	379,204	18.1	388,137	18.6
6.00 - 6.99	242,000	12.7	245,652	11.8	250,521	12.0
7.00 - 7.99	137,472	7.2	143,887	6.9	144,662	6.9
8.00 - 8.99	74,270	3.9	80,913	3.9	80,486	3.9
9.00 - 9.99	39,671	2.1	44,793	2.1	43,947	2.1
10.00 - 14.99	45,808	2.4	56,569	2.7	53,348	2.6
15.00 - 19.99	3,182	0.2	4,660	0.2	4,099	0.2
20.00 - 24.99	357	0.0	525	0.0	484	0.0
≥ 25.00	0	-	0	-	0	-
Total	1,912,257	100.0	2,090,182	100.0	2,090,182	100.0

From this table, it can be seen that only a small proportion of households had weights less than 1; in fact, fewer than 0.7% of households had weights this small. Overall, there was a higher proportion of small and large weights in 1996 than in 1991. Note that, in 1996, every weight size less than 3 has a larger percentage of the total than the corresponding group in 1991. In addition, every weight size of 9 or more also contains a larger percentage than the corresponding group in 1991. However, for the mid-range weights between 4 and 9, 1991 contains a larger percentage of households in this range. When the percentage of weights contained between 3 and 7 is examined from 1986 (Bankier *et al.*, 1992) and from 1991 to 1996, a steep decline is quite apparent. Whereas, in 1986, when the raking ratio estimation procedure (RREP) was used, 93.9% of the households had their weights in this range; in 1991, it dropped to 72.3% and in 1996, to 71.0% for the first production run and to 70.3% for the second production run.

2. Discrepancy Between Population Counts and Sample Estimates

One of the aims of the weighting procedure is to minimize the discrepancies between population counts and the corresponding sample estimates for the constraints. These discrepancies are the result of sampling variability and bias (see Chapter V). Even after the weighting procedure is completed, however, some discrepancies may remain. Discrepancies are measured by the difference between the sample estimate and the population count, expressed as a percentage of the population count, i.e.

$$\text{discrepancy} = \frac{\text{sample estimate} - \text{population count}}{\text{population count}} \times 100 \quad (2)$$

The numerator of the above expression (sample estimate – population count) is often referred to as the “**difference**”. Note that the population count is for sampled EAs only.

Table 6.4 shows the **differences** and **discrepancies** at the Canada level in 1996 for 32 basic demographic characteristics. This table displays the “**differences**” for three separate sample estimates: the initial weights, the weights from the first run and the weights from the second run. Note that the discrepancy has been rounded to two decimal places. The sample estimates and population counts are based on occupied private dwellings in sampled EAs.

The table clearly shows that the first run of the weighting system reduced the difference for most of the characteristics. For instance, the initial weights produced a discrepancy of 91,338 between the sample estimate and the population count for married persons, while the first run of the weighting system reduced this discrepancy to 10. The discrepancies for the younger age groups have also improved substantially. However, the weighting system did not improve the discrepancy for all characteristics. Indeed, for common-law status, the initial weights produced a difference of -1,404, which the weighting system increased to 46,646. Similar distortions also occurred in household size, especially for household size = 5 and household size = 6. Note that, for household size = 5, the initial weights produced a difference of 3,170, which the weighting system increased to a difference of 57,775. These results were the main factor in running the weighting system a second time.

As discussed above, the second run of the weighting system used common-law status as a constraint. In addition, the initial weights were adjusted to agree with the population counts of household size at the WA level. The results of the differences from the second run are also shown in Table 6.4. The main difficulties resulting from the first run were improved by the second run. For instance, the discrepancy for common-law was reduced from 46,646 to 2,415. Household size = 5 had its difference reduced from 57,775 to 27,879. However, some characteristics were made slightly worse by this second run. For example, persons aged 75 and over had a difference of -2,377 in the first run, but this increased to -9,207 with the second run.

Table 6.4 1996 Estimate/Population Discrepancies at the Canada Level

Constraint	Initial Weights	First Run		Second Run	
	Difference	Difference	Discrepancy	Difference	Discrepancy
Males	-22,868	-	0.00	15	0.00
Males over 14	-48,269	213	0.00	-276	0.00
Persons over 14	-29,285	3	0.00	3	0.00
Total households	1,060	-	0.00	-	0.00
Total population	23,117	-	0.00	-	0.00
Age 0-4	15,779	34	0.00	-208	-0.01
Age 5-9	18,705	168	0.01	-258	-0.01
Age 10-14	17,918	-205	-0.01	462	0.02
Age 15-19	4,709	1,890	0.10	1,853	0.10
Age 20-24	-24,353	1,501	0.08	803	0.04
Age 25-29	-17,831	-49	-0.00	105	0.01
Age 30-34	-3,979	248	0.01	361	0.02
Age 35-39	-3,924	493	0.02	320	0.01
Age 40-44	5,251	232	0.01	366	0.02
Age 45-49	9,004	1,743	0.09	971	0.05
Age 50-54	8,267	959	0.06	993	0.06
Age 55-59	-2,135	-201	-0.02	254	0.02
Age 60-64	2,533	-3,380	-0.29	3,847	0.33
Age 65-74 #	4,582	-1,056	-0.05	-662	-0.03
Age 75 and over	-11,408	-2,377	-0.19	-9,207	-0.74
Single persons	-37,340	239	0.00	115	0.00
Married persons	91,338	10	0.00	73	0.00
Widowed persons	-11,803	-1,338	-0.10	-1,387	-0.11
Divorced persons	-13,606	951	0.06	1,209	0.08
Separated persons	-5,472	137	0.02	-10	-0.00
Common-law = yes	-1,404	46,646	2.63	2,415	0.14
One-person households #	-	-	-	-4,750	-0.18
Two-person households	12,060	-344	-0.01	-3,331	-0.05
Three-person households	4,772	3,317	0.06	2,614	0.05
Four-person households	11,666	4,822	0.07	6,776	0.09
Five-person households	3,170	57,775	1.57	27,879	0.76
Six-or-more-person households #	-	-65,570	-1.38	-29,187	-0.52

= Constraint added for second run

3. Discarding Constraints

In the first run, a total of 29 constraints (see Appendix B) were used at the WA level, including five-year age groups, marital status, sex and household size. The constraint "Marital status – Separated" was not used because it was linearly dependent on the other marital status constraints. Similarly, the age constraint 60-64 and the household size 6 were not used because they were linearly dependent on other constraints. As discussed above, in order to retain more constraints, the weights in 1996 were allowed to be in the range of (0.25) instead of the range (1.25) used in 1991.

Table 6.5 below shows how often a constraint was discarded for the second run at the WA level and for what reasons. **Linearly dependent (LD)** implies that the constraint is redundant (six-or-more-person households is an example, since total households is a constraint; hence, one of the subgroups has to be redundant) and, as such, the removal will make no difference to the final estimates. **Small** implies that the constraint applies to less than 20 households, while **nearly linearly dependent (NLD)** and **outlier** implies the constraint causes the weights to be larger than 25 or less than 0.

This table shows how often a constraint was dropped; thus, out of a total of 5,941 WAs, the constraint age 0-4 was dropped 3,154 times. This implies that the potential for a discrepancy between the population count and the sample estimate using the regression weights exists. While there is very little discrepancy for this characteristic – Table 6.4 shows a discrepancy of -208 persons for ages 0-4 – this is because it was a **linearly dependent** constraint. However, for one-person households, the difference increases since it was being dropped for being **nearly linearly dependent**. Note that for one-person households, the constraint was dropped 4,600 times (4,583 times because of NLD) and, consequently, the discrepancy between the population count and the estimate using regression weights is -4,740. There is a clear link between the number of times a constraint is dropped and the resulting difference – provided the constraint is not dropped because of linearly dependence.

One group of characteristics does not follow this pattern. Note that the characteristic age from 0-4, 5-9 and 10-14 are all discarding with fairly high frequency. However, the constraints “persons over 15” and “total persons” are almost always retained. Hence, the larger grouping of age 0-14 is a redundant constraint which is why, in turn, the three age groups are dropped with such regularity for being linearly dependent.

Table 6.5 Frequency of Discarding WA Level Constraints in 1996

Characteristic	Small	LD	NLD	Outlier	Total
Males	0	0	0	1	1
Females*	-	-	-	-	-
Total population	0	0	0	0	0
Age 0-4	6	3,071	20	57	3,154
Age 5-9	30	709	77	135	951
Age 10-14	35	2,110	33	61	2,239
Age 15-19	6	514	27	96	643
Age 20-24	1	216	133	119	469
Age 25-29	1	347	108	82	538
Age 30-34	1	29	23	42	95
Age 35-39	1	0	6	31	38
Age 40-44	1	3	13	45	62
Age 45-49	1	4	9	50	64
Age 50-54	2	157	67	83	309
Age 55-59	2	636	213	147	998
Age 60-64	3	1,122	973	128	2,226
Age 65-74	4	3	214	81	302
Age 75 and over	36	2,864	100	60	3,060
Single persons	0	0	1	3	4
Married persons	0	0	0	4	4
Widowed persons	2	0	174	345	521
Divorced persons	1	1	213	252	467
Separated persons*	-	-	-	-	-
Common-law = yes	23	0	1	272	296
One-person households	1	12	4,583	4	4,600
Two-person households	0	0	1,154	12	1,166
Three-person households	2	22	189	47	260
Four-person households	23	145	52	37	257
Five-person households	193	997	865	92	2,147
Six-or-more-person households*	-	-	-	-	-
Males aged ≥ 15	0	1	136	3	140
Persons aged ≥ 15	0	0	1	0	1
Total households	0	0	0	0	0

* Indicates the characteristic was not used as a constraint because it was redundant.

4. Evaluation at Various Geographic Levels

A study was done which compared the absolute differences between sample estimates and population counts for 31 characteristics in 1996 (29 of which were constraints) and 1986 for various geographic levels. The 31 characteristics that were part of this study of absolute differences are listed in Appendix B. The results of the study are summarized in Table 6.6 below. The table contains the percentage of characteristics that had an "R value" within a certain range for the six geographic levels shown in the table. An R value is a ratio between 1996 and 1986 differences obtained from the following equation:

$$R = 100 * \frac{\sum_{i=1}^{N_{96}} |\hat{X}^{96}_i - X^{96}_i| / \sum_{i=1}^{N_{86}} |X^{86}_i|}{\sum_{i=1}^{N_{96}} |\hat{X}^{96}_i - X^{96}_i| / \sum_{i=1}^{N_{86}} |X^{86}_i|} \quad (3)$$

where X^{96} and X^{86} are respectively the 1996 and 1986 population counts for a characteristic. The sample estimate in 1996 based on GLSEP weights is \hat{X}^{96} while the sample estimate in 1986 based on RREP weights is \hat{X}^{86} . R values were calculated for each of the six geographic levels (EA, WA, CSD, CD, Province and Canada). The sum of the absolute values of the population/estimate differences were calculated, where N_{96} equals the number of areas for the particular geographic level in 1996 and N_{86} equals the number of areas for the particular geographic level in 1986. An R value in the range of 95 to 105 means that the 1996 estimation system and 1986 estimation system performed equally well. An R value less than 95 means that the 1996 system performed better than the 1986 system for the characteristic at the particular geographic level, while an R value greater than 105 means that it did worse. The results for 1991, when the comparison was made with 62 characteristics (49 of which were constraints), are also shown.

Table 6.6 Percentage of the Characteristics With R Values Falling in Certain Ranges

	R Value	EA	WA	CSD	CD	Province	Canada
1996 vs 1986	< 95	94	64	94	64	52	49
	95-105	0	15	0	15	21	18
	>105	6	21	6	21	27	33
1991 vs 1986	< 95	87	58	81	47	31	29
	95-105	11	11	8	18	14	10
	>105	2	31	11	35	55	61

The 1996 estimation system was effective at reducing the population/estimate differences at all levels compared to the 1986 estimation system. Note that, in 1991, the estimation system improved the differences at the smaller geographic levels – from EA up to CD – but that, at the province and Canada level, the 1986 estimation system had smaller differences.

VII. Sample Estimate and Population Count Consistency

In order for the GLSEP to work well, some of the constraints had to be discarded within each WA before the weights could be calculated. Consequently, many important characteristics were discarded in a number of WAs. As a result, the level of agreement (consistency) between sample estimates and population counts for these characteristics was reduced. Furthermore, many geographic areas of interest do not always consist of complete WAs. Consequently, in these areas the consistency for all characteristics depends on how close the areas come to consisting of complete WAs.

The consistency study examined the discrepancies between sample estimates and population counts (expressed as percentages of the population counts) for the same basic set of 31 characteristics as the Sampling Bias Study (see Appendix B) for the following geographic areas:

- (a) census divisions;
- (b) census subdivisions;
- (c) census tracts;
- (d) enumeration areas.

As in Chapter VI (Subsection B.2), the discrepancies between sample estimates and population counts were calculated as:

$$\text{discrepancy} = \frac{\text{sample estimate} - \text{population count}}{\text{population count}} \times 100$$

A. Census Divisions (CDs)

The percentiles in Table 7.1 summarize the level of consistency for all sampled CDs in Canada for a wide variety of basic characteristics with a population count⁶ greater than 50. Generally, the discrepancies produced for characteristics with population counts ≤ 50 for most geographic areas were found to be relatively large (either positive or negative). Therefore, it was decided to not include geographic areas where the characteristic count was less than or equal to 50, because a few of these areas could significantly alter the percentiles of discrepancies in the tables in this chapter. This alteration would occur if many of these areas had either relatively large positive discrepancies or relatively large negative discrepancies.

In Table 7.1, for each characteristic, N% of the CDs had discrepancies that were less than the Nth percentile while $100 - N\%$ of the CDs had discrepancies that were greater than the Nth percentile. Thus, the discrepancy was between the 10th and 90th percentiles for 80% of the CDs, and the discrepancy was between the 25th and 75th percentiles for 50% of the CDs, etc. For example, the discrepancy for age 0-4 was between -0.44% and 0.05% for

⁶ The population count here refers to that of the characteristic. For example, the level of consistency for age 0-4 is summarized for all CDs in which there were more than 50 people in the age group 0-4. The same definition applies to Tables 7.1, 7.2, 7.3 and 7.4.

80% of the CDs. A pattern that is symmetric about 0 implies that the difference between the sample estimates and the population counts is spread in an even fashion. However, a non-symmetric distribution implies that sample counts are not evenly spread out between positive and negative differences.

All CDs consist of complete WAs. Thus the characteristics which were constraints in 1996 and which were rarely or never discarded in a WA had nearly perfect consistency at the CD level⁷. These characteristics include total number of persons, total number of females, number of single persons, and so on. The level of consistency for most of the remaining characteristics, while not perfect, was still quite good, except for those characteristics which represent only a small percentage of the population in most CDs, such as the number of separated persons or the number of six-or-more-person households. A general relationship does exist between the discrepancies and the population counts for all characteristics, in that the consistency improves as the population count for the CD increases.

Table 7.1 also shows the corresponding percentiles of the 1991 discrepancies for CDs. Tables 7.2, 7.3 and 7.4 which follow also contain the 1991 data for the other geographic levels. When comparing the 1991 and 1996 data, the 1996 discrepancies at the CD level are the same as or significantly smaller than the 1991 discrepancies for all of the characteristics in Table 7.1.

The sizes of the discrepancies at the CD level are quite small compared to the discrepancies at the smaller geographic levels that are studied in the following sections.

B. Census Subdivisions (CSDs)

Table 7.2 summarizes the level of consistency between sample estimates and population counts for all sampled CSDs in Canada with a population count for the characteristic greater than 50. It includes the same characteristics as Table 7.1. CSDs do not always consist uniquely of complete WAs. They are also much smaller on average than CDs. Consequently, the consistency was not as good for CSDs as for CDs. In general, as with CDs, the consistency improved as the population count for the CSD increased, for all characteristics. In comparison to 1991, the range of discrepancies between the 10th and 90th percentiles is smaller in 1996 for almost all of the characteristics. This is also true for the 25th to 75th percentile range.

C. Census Tracts (CTs)

Table 7.3 summarizes the level of consistency for all sampled CTs in Canada. As with Tables 7.1 and 7.2, this table only includes CTs where population counts for the characteristic were greater than 50. CTs have larger populations on average than CSDs. The improvement in the level of consistency at the CT level between 1991 and 1996 is very noticeable. The 25th and 75th percentiles for 1996 are equal to zero for almost all characteristics, contrary to the ranges found in 1991. Also, the ranges between the 10th and 90th percentiles are much smaller in 1996.

New specifications were used to create WAs in 1996. These specifications created fewer CT boundary crossings than in 1991. This change could help explain the improvements noted in 1996.

⁷ Even for characteristics with perfect consistency, published tabulations of basic characteristics based on sample data will not agree exactly with tabulations of the same characteristics based on 100% data. This is because those residents of collective dwellings who were not asked the sample questions (see Chapter II, Section B) are included in tabulations based on 100% data, but are excluded from tabulations based on sample data.

D. Enumeration Areas (EAs)

EAs are the components of WAs, and WAs are the lowest level at which sample estimates are forced to agree with population counts for most characteristics. EAs are also the components of higher geographic levels (CDs, CSDs, CTs, etc.) and a number of the WAs are, as Table 7.3 shows, components of these higher levels. Consequently, the consistency at the EA level cannot be expected to be as good as it would be at the higher geographic levels that have been studied. Table 7.4 confirms this as it shows that, for most characteristics studied in sampled EAs with a population count for the characteristic greater than 50, the discrepancies are larger than the discrepancies for the geographic levels studied earlier. This is the case in both 1996 and 1991. In comparison to the 1991 discrepancies at the 10th and 90th percentiles (and at the 25th and 75th percentiles), the 1996 discrepancies are lower for most of the characteristics studied. The values for one-person households, however, are very similar *for* 1991 and 1996.

E. Impact of the Changes to the Weighting Procedure in 1996

There are a number of possible explanations for the improvement in consistency found in 1996 compared to 1991:

- (a) Fewer constraints were used in 1996. This resulted in fewer constraints being dropped, and allowed the set of constraints used in a particular WA to be very similar to the sets used in other WAs. The consistency between the estimate and the count for a given characteristic is thus more stable among WAs.
- (b) In 1996, the weights calculated by the procedure could be smaller than one. This fact also contributed to a reduction in the number of constraints dropped.
- (c) Changes in WA formation reduced the number of higher geographic area boundary crossings. Some larger geographic areas now consist of more complete WAs. This contributes to reducing the discrepancies in these areas.

In conclusion, it appears that, for smaller geographic areas, the changes introduced to the weighting procedure for 1996 generally yielded better estimates than in 1991 in the sense that, for areas of the same geographic level, the discrepancy for a given characteristic is generally closer to 0 than for the same characteristic in 1991.

Table 7.1 Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CDs – 1996 and 1991
Censuses

Characteristics Studied	1996 Percentiles					1991 Percentiles				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Person Characteristics										
Males	0	0	0	0	0	0	0	0	0	0
Females	0	0	0	0	0	0	0	0	0	0
Total population	0	0	0	0	0	0	0	0	0	0
Age 0-4	-0.44	0	0	0	0.05	-2.86	-1.07	0	1.06	2.38
Age 5-9	-0.10	0	0	0	0.21	-2.14	-1.05	0	0.63	2.06
Age 10-14	0	0	0	0	0.04	-1.80	-0.56	0	0.99	2.36
Age 15-19	0	0	0	0	0.43	-1.87	0.51	0.55	1.76	3.15
Age 20-24	-0.89	0	0	0.09	1.17	-3.71	-0.95	0.32	2.35	4.14
Age 25-29	-0.95	0	0	0	1.32	-3.07	-1.40	-0.20	0.20	1.78
Age 30-34	0	0	0	0	0	-1.67	-0.39	0	0.62	2.11
Age 35-39	0	0	0	0	0	-3.14	-0.84	0	0.59	2.26
Age 40-44	0	0	0	0	0	-2.33	-0.69	0	1.02	3.04
Age 45-49	0	0	0	0	0	-2.95	-1.14	0	1.80	4.26
Age 50-54	-0.17	0	0	0	0.21	-4.96	-2.14	0.13	2.02	5.03
Age 55-59	-1.05	0	0	0	0.85	-6.13	-2.33	0	1.60	4.21
Age 60-64	-1.51	-0.01	0	0.92	2.52	-3.69	-1.75	0.07	1.93	4.88
Age 65-74	-0.25	0	0	0	0.06	-2.28	-1.08	0	0.59	2.03
Age 75 and over	-3.70	-1.94	-0.17	0	0.81	-7.87	-3.66	-1.07	0.67	4.65
Single persons	0	0	0	0	0	-0.10	0	0	0	0.12
Married persons	0	0	0	0	0	0	0	0	0.08	0.33
Widowed persons	-1.53	-0.33	0	0.09	1.60	-4.22	-2.26	-0.55	0.57	2.27
Divorced persons	-1	0	0	0.34	2.04	-4.47	-1.82	-0.14	1.88	4.73
Separated persons	-7.29	-2.25	0	1.15	4.30	-9.33	-3.96	0.44	4.90	11.12
Common-law = yes ¹	-0.86	0	0	0	1.36	-	-	-	-	-
Household Characteristics										
One-person households	-0.57	-0.36	-0.20	-0.10	0.10	-2.54	-1.44	-0.60	0.03	0.73
Two-person households	-0.10	0	0	0	0.01	-0.51	0	0	0.22	0.87
Three-person households	0	0	0	0	0.10	-2.43	-0.92	0.21	1.23	3.17
Four-person households	0	0	0	0	0.16	-1.66	-0.57	0.04	0.92	2.15
Five-person households	-0.59	0	0.60	1.81	3.09	-3.83	-0.59	1.56	4.19	6.79
Six-or-more-person households	-6.55	-3.39	-0.84	0.99	2.65	-14.50	-8.69	-4.20	0.18	5.21

¹ Common-law = yes was not used as a constraint in 1991. Thus, comparison between 1991 and 1996 for this characteristic does not make sense.

Table 7.2 Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CSDs – 1996 and 1991 Censuses

Characteristics Studied	1996 Percentiles					1991 Percentiles				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Person Characteristics										
Males	-4.27	-1.22	0	1.30	4.17	-5.47	-1.69	0	1.62	5.05
Females	-4.49	-1.33	0	1.22	4.27	-5.49	-1.83	0	1.83	5.56
Total population	-2.11	0	0	0	1.94	-3.36	-0.37	0	0.26	3.28
Age 0-4	-16.40	-4.20	0	3.53	15.47	-19.40	-7.14	0	6.23	18.50
Age 5-9	-13.20	-3.45	0	3.66	13.98	-16.50	-6.11	0	5.73	17.01
Age 10-14	-13.00	-4.19	0	3.91	13.02	-17.00	-6.15	0	7.06	18.01
Age 15-19	-12.90	-3.20	0	3.87	13.64	-16.80	-6.03	0	6.92	19.04
Age 20-24	-14.10	-3.97	0	3.47	14.06	-21.00	-7.71	0	7.94	21.04
Age 25-29	-14.20	-3.27	0	4.03	15.15	-20.30	-7.54	0	5.96	19.29
Age 30-34	-12.70	-3.40	0	3.99	12.94	-17.90	-6.21	0	6.02	17.39
Age 35-39	-13.00	-3.43	0	3.70	13.23	-19.20	-6.66	0	5.86	18.71
Age 40-44	-13.50	-3.70	0	3.49	12.81	-19.20	-6.43	0	7.42	19.92
Age 45-49	-13.00	-3.27	0	3.90	13.81	-19.20	-7.37	0	8.45	22.15
Age 50-54	-14.60	-3.46	0	3.27	13.84	-23.10	-9.06	0	8.62	21.29
Age 55-59	-14.80	-2.95	0	4.00	15.20	-22.30	-8.28	0	8.34	22.23
Age 60-64	-15.70	-3.73	0	4.79	15.54	-22.30	-9.43	0	9.14	22.91
Age 65-74	-14.80	-4.25	0	3.91	13.33	-21.20	-8.49	0	6.72	19.36
Age 75 and over	-18.40	-5.91	0	4.14	15.11	-26.20	-12.00	-1.31	7.74	22.02
Single persons	-5.46	-1.61	0	1.83	5.46	-7.26	-2.34	0	2.33	7.14
Married persons	-5.83	-1.83	0	1.73	5.64	-6.33	-1.99	0	2.31	6.98
Widowed persons	-14.40	-4.13	0	3.55	14.45	-18.90	-8.35	0	6.03	17.18
Divorced persons	-14.90	-2.73	0	3.20	15.10	-19.60	-7.28	0	7.72	19.43
Separated persons	-11.10	-0.82	0	0.88	10.51	-20.50	-7.85	0	8.55	22.51
Common-law = yes	-17.40	-4.84	0	3.96	16.12	-	-	-	-	-
Household Characteristics										
One-person households	-11.10	-3.16	-0.14	2.87	10.48	-11.90	-5.07	-0.54	3.51	10.46
Two-person households	-10.80	-3.29	0	3.15	10.09	-11.10	-3.71	0	4.02	11.56
Three-person households	-10.60	-2.56	0	2.82	11.37	-15.30	-5.54	0	6.18	17.21
Four-person households	-10.20	-2.93	0	2.60	10.11	-14.80	-5.03	0	4.44	13.50
Five-person households	-7.89	-0.91	0	3.69	11.42	-14.00	-4.94	0.40	8.11	19.45
Six-or-more-person households	-9.36	-3.29	-0.28	2.12	5.84	-20.80	-10.40	-3.38	4.13	12.07

Table 7.3 Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for CTs – 1996 and 1991
Censuses

Characteristics Studied	1996 Percentiles					1991 Percentiles				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Person Characteristics										
Males	-0.20	0	0	0	0.26	-0.90	0	0	0	0.87
Females	-0.20	0	0	0	0.25	-0.90	0	0	0	0.79
Total population	0	0	0	0	0	-0.20	0	0	0	0.24
Age 0-4	-1.40	0	0	0	1.80	-11.00	-1.20	0	1.03	9.77
Age 5-9	-1.90	0	0	0	1.72	-11.00	-1.60	0	0.94	10.20
Age 10-14	-1.90	0	0	0	1.72	-11.00	-1.30	0	2.23	13.00
Age 15-19	-3.10	0	0	0	3.62	-11.00	-2.70	0	3.93	11.60
Age 20-24	-2.10	0	0	0	2.44	-10.00	-1.50	0	1.90	10.80
Age 25-29	-1.90	0	0	0	1.89	-10.00	-1.60	0	0.28	7.31
Age 30-34	-1.10	0	0	0	1.28	-7.80	-0.60	0	0.29	7.46
Age 35-39	-1.00	0	0	0	1.16	-9.30	-1.30	0	0	7.88
Age 40-44	-1.20	0	0	0	1.55	-10.00	-1.40	0	1.02	9.05
Age 45-49	-1.10	0	0	0	1.64	-13.00	-3.10	0	3.06	12.00
Age 50-54	-1.80	0	0	0	2.48	-15.00	-5.00	0	5.84	16.00
Age 55-59	-3.60	0	0	0	3.62	-16.00	-5.00	0	5.84	17.30
Age 60-64	-7.00	0	0	0	9.89	-17.00	-6.50	0	6.54	18.60
Age 65-74	-2.70	0	0	0	2.34	-14.00	-3.00	0	1.61	12.40
Age 75 and over	-14.00	0	0	0	7.31	-22.00	-8.90	0	8.23	21.00
Single persons	-0.30	0	0	0	0.31	-1.20	0	0	0	1.27
Married persons	-0.40	0	0	0	0.35	-1.40	0	0	0	1.59
Widowed persons	-4.40	0	0	0	3.36	-15.00	-5.30	0	3.87	14.90
Divorced persons	-2.20	0	0	0	2.97	-16.00	-5.20	0	4.08	14.20
Separated persons	-4.60	0	0	0	5.53	-24.00	-8.80	0	8.86	25.50
Common-law = yes	-2.50	0	0	0	3.39	-	-	-	-	-
Household Characteristics										
One-person households	-2.80	-0.80	0	0.29	2.06	-7.60	-3.20	-0.40	1.38	5.39
Two-person households	-1.50	0	0	0	1.11	-3.90	-0.30	0	0.40	4.23
Three-person households	-1.60	0	0	0	1.79	-8.90	-2.30	0	3.32	9.95
Four-person households	-0.90	0	0	0	2.01	-7.60	-0.90	0	2.19	9.90
Five-person households	-3.60	0	0	0.12	8.01	-16.00	-4.80	0	7.47	18.50
Six-or-more-person households	-6.00	-2.30	1.04	3.78	7.23	-24.00	-11.00	-0.50	7.11	16.30

Table 7.4 Percentiles of Sample Estimates and Population Count Discrepancies (as a Percentage of the Population Count) for EAs – 1996 and 1991 Censuses

Characteristics Studied	1996 Percentiles					1991 Percentiles				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
Person Characteristics										
Males	-4.97	-2.05	0.02	2.14	4.80	-6.25	-2.59	0.07	2.64	6.06
Females	-4.91	-2.08	0	1.96	4.77	-6.29	-2.66	-0.10	2.55	6.24
Total population	-1.94	0	0	0	1.81	-3.56	0	0	0	3.35
Age 0-4	-21.40	-10.30	0	10.60	21.78	-26.40	-12.70	-0.22	12.50	26.49
Age 5-9	-17.30	-8.68	-0.10	8.63	17.87	-24.20	-11.50	0.03	11.23	24.00
Age 10-14	-19.20	-9.33	-0.15	9.63	20.01	-25.40	-12.20	0.07	12.23	25.28
Age 15-19	-16.70	-8.26	0.11	8.94	17.86	-25.40	-12.10	0.37	13.15	27.12
Age 20-24	-18.30	-9.30	-0.12	9.39	19.20	-27.90	-13.60	-0.23	13.51	28.48
Age 25-29	-18.40	-9.44	-0.32	9.42	19.36	-26.10	-12.60	-0.51	11.45	25.23
Age 30-34	-17.10	-8.73	-0.14	8.56	17.72	-24.40	-11.60	-0.23	11.43	24.63
Age 35-39	-16.90	-8.53	-0.30	8.22	17.14	-25.10	-12.40	-0.68	11.28	24.60
Age 40-44	-17.20	-9.02	-0.34	8.52	17.92	-26.30	-12.80	-0.37	11.93	26.53
Age 45-49	-17.30	-9.02	-0.40	8.60	18.39	-27.40	-13.10	0	13.65	29.16
Age 50-54	-19.00	-9.59	-0.54	9.37	19.54	-30.00	-15.20	-0.10	14.85	31.77
Age 55-59	-19.90	-10.60	-0.74	9.96	20.66	-29.60	-15.00	0.29	15.76	31.40
Age 60-64	-20.50	-10.30	0.06	10.82	22.07	-30.30	-15.40	-0.50	16.01	31.98
Age 65-74	-17.30	-9.12	-0.40	8.60	17.73	-26.60	-12.60	-0.63	11.72	27.41
Age 75 and over	-18.20	-8.72	-0.06	8.78	18.10	-27.50	-12.70	-0.32	12.50	27.83
Single persons	-6.74	-2.76	-0.39	2.84	6.22	-8.45	-3.43	0.04	3.42	8.03
Married persons	-7.85	-3.39	0.06	3.41	8.08	-8.78	-3.55	0.07	3.76	8.92
Widowed persons	-18.00	-9.39	0.10	8.69	17.55	-23.50	-11.50	-0.60	10.68	22.99
Divorced persons	-18.90	-9.61	0	9.86	19.77	-29.70	-15.20	-0.93	14.50	29.84
Separated persons	-33.10	-20.80	-0.36	11.74	29.78	-28.90	-11.70	3.21	17.38	32.19
Common-law = yes	-6.00	0	0	0	6.66	-	-	-	-	-
Household Characteristics										
One-person households	-13.10	-6.01	0.01	6.00	12.62	-13.50	-6.67	-0.35	5.69	12.38
Two-person households	-13.80	-6.96	-0.23	6.61	13.67	-14.40	-7.08	0	7.06	15.11
Three-person households	-16.00	-8.40	-0.17	7.88	16.21	-22.30	-11.40	-0.12	11.25	23.48
Four-person households	-13.90	-7.04	-0.19	7.07	14.20	-17.60	-8.41	0	8.49	18.45
Five-person households	-16.40	-8.23	-0.10	8.46	16.68	-22.20	-11.00	1.09	13.92	26.33
Six-or-more-person households	-13.30	-5.98	2.01	10.03	18.95	-22.60	-9.30	2.48	14.81	26.63

VIII. Sampling Variance

A sampling error can be divided into two components: variance and bias. The variance measures the variability of the estimate about its average value in hypothetical repetitions of the survey process, while the bias is defined as the difference between the average value of the estimate in hypothetical repetitions and the true value being estimated. The mean square error (MSE) measures the variability of the estimate about the true value in hypothetical repetitions of the survey process. It can be shown that the MSE equals the variance plus the square of the bias. The MSE measures most accurately how far the estimate is from the true population value on average. If the bias is small relative to the variance, however, the variance is a good approximation of the MSE. There is evidence, however, that the bias accumulates as census estimates for progressively larger geographic areas are produced. Thus, the bias can be insignificant for small geographic areas but become large relative to the variance for large geographic areas. Because of this, the variance can be much smaller than the MSE for large geographic areas. The variance of an estimate can be estimated from the sample, but the bias of an estimate cannot. This means that it is not possible to accurately estimate the MSE from the sample unless the bias is small relative to the variance.

In previous censuses, a study to provide estimates of the sampling variance was carried out. A few results from the 1986 study are provided in Section A (for more information, see the *User's Guide to the Quality of 1986 Census Weighting: Sampling and Weighting*). Because at larger geographic levels the bias is felt to be the dominant term in the MSE (see Chapter V), calculating the sampling variance does not provide an accurate estimate of the MSE for large geographic areas; hence, it was decided not to repeat this study for the 1996 Census. A discussion is given in Section B, however, of what impact the estimation methodology used in the 1996 Census had on the sampling variance compared to the 1986 Census estimation methodology.

A. 1986 Census Sampling Variance Study

Chapter V presented results of the Sampling Bias Study, describing the nature and extent of bias in the census sample prior to weighting. Chapters VI and VII presented results on the sampling bias following the application of the weighting procedure. Even with a perfectly unbiased sampling method, the results would still be subject to variance, simply because the estimates are based only on a sample. The variance may be estimated using the data collected by the sample survey.⁸ The 1986 Sampling Variance Study was carried out to estimate the effect of the sampling and estimation procedures on those census figures that are based on sample data.

On the basis of the 2B sample data, thousands of tables are produced by Statistics Canada. Conceptually, the estimated sampling variance, which is a measurement of precision, can be associated with every estimate calculated in these tables. This measurement takes into account both the sample design and the estimation method. In practice, however, it cannot be calculated for every census estimate because of high data processing costs. Sampling variance is thus estimated for only a subset of census estimates. From this, the combined effect of the sample design and the estimation method on the sampling variance can be estimated. Simple estimates of sampling variance, which are inexpensive to calculate, can then be adjusted for this impact to produce estimates of sampling variance for any census estimates.

Table 8.1 gives the non-adjusted (simple) standard errors of census sample estimates. The figures in this table were determined by assuming that 1 in 5 simple random sampling and simple weighting by 5 was used. The standard errors are expressed in Table 8.1 as a function of the size of both the census estimate and the geographic area. For example, for an estimate of 50 persons in a geographic area with a total of 500 persons, the non-adjusted standard error is 15.

⁸ Unfortunately, the sampling variance does not provide any indication of the extent of non-sampling error.

Standard errors are given in Table 8.1 for only a limited number of values for the estimated total and the total number of persons, households, dwellings or families in the area. The following formula may be used to calculate the non-adjusted standard errors for any estimated total for an area of any size:

$$NASE = \sqrt{\frac{4E(N-E)}{N}} \quad (4)$$

where NASE is the non-adjusted standard error, E is the estimated total and N is the total number of persons, households, dwellings or families in the area. For example, for an estimated total of 750 persons in an area with a total of 9,000 persons, the non-adjusted standard error would be:

$$\sqrt{\frac{4(750)(9,000 - 750)}{9,000}} = 52$$

The 1986 Sampling Variance Study provides adjustment factors⁹ by which the non-adjusted standard errors should be multiplied to adjust for the combined effect of the sample design and the estimation procedure. To calculate these adjustment factors, a sample of 401 WAs (out of a total of 5,941 WAs) was selected. The sample was allocated among the ten provinces¹⁰ in such a way as to obtain good estimates of the sampling variance at the provincial level without greatly sacrificing the quality of the estimates at the national level. For each WA in the sample, estimates of the sampling variances for **raking ratio estimates** were calculated for different categories of all the characteristics given in **Table 9 of the 1986 Census User's Guide**; this table is included as Appendix D. (**IMPORTANT NOTE:** These factors were calculated from the 1986 Census, which used the RREP; using this table assumes the adjustment effects are the same for the 1996 Census, even though a different estimation procedure was used.) The estimates of sampling variance at the provincial and national levels were obtained by weighting up the WA level estimates. The adjustment factors for each category of each characteristic were calculated by dividing the square roots of the WA level estimates by the non-adjusted standard errors. Adjustment factors were calculated at the provincial and national levels for each characteristic by averaging the adjustment factors for all of its categories. For further information on how these adjustment factors were calculated, see Béland, 1990.

To estimate the standard error for a given census sample estimate, the adjustment factor applying to the characteristic was determined from Appendix D. The adjustment factor at the national or provincial level for sample characteristics was generally in the range 0.40 to 1.60. Then this factor was multiplied by the non-adjusted standard error selected in Table 8.1.

The following example illustrates how to calculate the adjusted standard errors. Suppose the estimate of interest is the immigrant population in Ontario. The 1986 estimate for this characteristic was 2,081,200. The 1986 Census count for the population of Ontario was 9,001,170. Using equation (4) – which calculates the non-adjusted standard error – results in an estimate of 2,530. From Appendix D, the provincial level adjustment factor for the characteristic “immigrant” is 1.12. Consequently, the adjusted standard error for this estimate is $2,530 \times 1.12 = 2,834$.

⁹ The squares of the adjustment factors are commonly known as “design effects”.

¹⁰ The Yukon Territory and Northwest Territories were grouped with British Columbia.

A second example, however, casts doubt on the accuracy of these adjusted standard errors as estimates of the square root of the MSE. The estimated number of persons in the 1986 Census with “Marital status – Married” who lived in private dwellings in sampled EAs was 11,771,126. The number of persons in the 1986 Census who lived in private dwellings in sampled EAs was 24,369,559. Applying equation (4) generates a non-adjusted standard error of 4,934. From Appendix D, the national level adjustment factor for the characteristic “Married” is 0.25. Consequently, the adjusted standard error for this estimate is $4,934 \times 0.25 = 1,233$. Because marital status is a basic characteristic, however, it is known that the population count of the number of persons in the 1986 Census with “Marital status – Married” who lived in private dwellings in sampled EAs was 11,778,842. The difference between the estimate and the population count is -7,716. The ratio of this difference to the adjusted standard error is $-7,716/1,233 = -6.25$. A 95% confidence interval for an estimate would normally be defined as plus or minus two times the adjusted standard error. The fact that the ratio of the difference to the standard error is -6.25 suggests that the adjusted standard error of 1,233 is an underestimate of the square root of the MSE.

B. Sampling Variance and Bias With Generalized Least Squares Estimation

In Bankier, Rathwell and Majkowski, 1992, the coefficients of variation (CVs) of the GLSEP for some sample characteristics were compared to the corresponding CVs of the RREP. In both cases, the 1986 Census data were used. The CV of an estimate is the square root of the estimated variance expressed as a percentage of the estimate. For 79 WAs, the estimated CVs were calculated for estimates of 507 EA level and 642 WA level sample characteristics (all of which applied to at least an estimated 60 households in the population). The WA level and EA level estimates were each classified into small estimates (less than or equal to the median value of the estimates) and large estimates (greater than the median value of the estimates). It was found that the median value for the CVs for large WA estimates was 5% for the GLSEP while it was 6% for the RREP. The median value for the CVs for small WA estimates was 13% for the GLSEP while it was 15% for the RREP. The median value for the CVs for large EA estimates was 10% for the GLSEP while it was 12.5% for the RREP. The median value for the CVs for small EA estimates was 15% for the GLSEP while it was 17.5% for the RREP. Thus, there was some reduction in the CVs for the GLSEP compared to the RREP at both the EA and WA levels. Because the variances at higher geographic levels are just the sum of the variances at the WA level, these reductions in the CVs should also hold at higher geographic levels.

Chapter V indicated that the census sample has small but significant biases. These biases are insignificant compared to the sampling variance at the WA level. For higher geographic levels, however, the bias for a characteristic can accumulate if the bias almost always results in overestimates or underestimates. It appears that the effect of the bias is more significant with the GLSEP than with the RREP. This can be seen from Table 6.5 in Chapter VI where the GLSEP has smaller population/estimate differences than the RREP for smaller geographic areas. However, as the geographic area grows, the improvement is not as large, and in 1991 the RREP was superior to the GLSEP at the provincial level. Besides bias introduced by sampling and processing, Bankier, Rathwell and Majkowski, 1992, show in a Monte Carlo study that the GLSEP estimator is biased, though the relative bias is less than 1% for 50% of the characteristics studied. More serious, however, is the fact that the estimated variance of GLSEP estimators has a median relative bias of -25% at the WA level. Thus, they tend to underestimate the true variance. The RREP estimators may suffer from similar biases, but no study of them has been done.

Table 8.1 Non-adjusted Estimates of Standard Errors of Sample Estimates

Estimated Total in the Area	Estimated Total Number of Persons, Households and Dwellings								
	500	1,000	2,500	5,000	10,000	25,000	50,000	100,000	250,000
50	15	15	15	15	15	15	15	15	15
100	18	19	20	20	20	20	20	20	20
250	22	25	30	30	30	30	30	30	30
500	0	30	40	40	45	45	45	45	45
1,000		0	50	55	60	60	65	65	65
2,500			0	70	85	95	95	100	100
5,000				0	100	130	130	140	140
10,000					0	150	180	190	200
25,000						0	220	270	300
50,000							0	320	400
100,000								0	490
250,000									0

Estimated Total in the Area	Estimated Total Number of Persons, Households and Dwellings					
	500,000	1,000,000	2,500,000	5,000,000	10,000,000	25,000,000
50	15	15	15	15	15	15
100	20	20	20	20	20	20
250	30	30	30	30	30	30
500	45	45	45	45	45	45
1,000	65	65	65	65	65	65
2,500	100	100	100	100	100	100
5,000	140	140	140	140	140	140
10,000	200	200	200	200	200	200
25,000	310	310	310	320	320	320
50,000	420	440	440	440	450	450
100,000	570	600	620	630	630	630
250,000	710	870	950	970	990	990
500,000	0	1,000	1,260	1,340	1,380	1,400
1,000,000		0	1,550	1,790	1,900	1,960
2,500,000			0	2,240	2,740	3,000
5,000,000				0	3,160	4,000
10,000,000					0	4,900

IX. Conclusion

Sampling is now an accepted and integral part of census-taking. Its use can lead to substantial reductions in costs and respondent burden associated with a census, or alternatively, can allow the scope of a census to be broadened at the same cost. The price paid for these advantages is the introduction of sampling error to census figures that are based on the sample. The effect of sampling is most important for small census figures, whether they are counts for rare categories at the national or provincial level or counts for categories in small geographic areas. It should be noted that response errors and processing errors also contribute to the overall error of census figures, and that it is the same small census figures that are particularly susceptible to the effects of these non-sampling errors. Therefore, even with a 100% census, many small figures would be of limited reliability. As a general rule of thumb for the 1996 Census, figures of size 50 or less that are based on sample data are of very low reliability, while figures up to size 500 tend to have standard errors in excess of 10% of their size.

For many of the characteristics, a certain amount of bias was detected in the sample. A small portion of the bias was found to have been introduced during data processing and edit and imputation. The remaining bias must have been due to one or more factors such as non-response bias, response bias, or the selection of a biased sample by the CRs. The procedures for weighting the sample data up to the population level were carried out successfully, and generally achieved the levels of sample estimate and population count consistency anticipated. While the consistency that was achieved at the provincial and Canada levels was better than in 1991, it is still lower than what might otherwise be expected given the improved consistency for smaller geographic levels that has been achieved. This is probably the result of the accumulation of small biases in the sample summed over many areas.

The census estimation methodology will be reassessed for the 2001 Census to see if it is possible to improve sample estimate and population count consistency at the provincial and Canada levels while maintaining good consistency at the EA level. Doing this should also allow more reliable estimates of the mean square error of the census estimates to be produced.

APPENDICES

Appendix A – Glossary of Terms

The definitions of census terms, variables and concepts are presented here as they appear in the *1996 Census Dictionary* (Catalogue No. 92-351-XPE). Users should refer to the *1996 Census Dictionary* for full definitions and additional remarks related to any concepts, such as information on direct and derived variables and their respective universe.

Census division (CD): Refers to the general term applied to areas established by provincial law which are intermediate geographic areas between the municipality (census subdivision) and the province level. Census divisions represent counties, regional districts, regional municipalities and other types of provincially legislated areas.

Census subdivision (CSD): Refers to the general term applying to municipalities (as determined by provincial legislation) or their equivalent (for example, Indian reserves, Indian settlements and unorganized territories).

Census tract (CT): Small geographic units representing urban or rural neighbourhood-like communities created in census metropolitan areas and census agglomerations (with an urban core population of 50,000 or more at the previous census).

Enumeration area (EA): Refers to the geographic area canvassed by one census representative. It is the smallest standard geographic area for which census data are reported. All the territory of Canada is covered by EAs.

Household: Refers to a person or a group of persons (other than foreign residents), who occupy the same dwelling and do not have a usual place of residence elsewhere in Canada. It may consist of a family group (census family) with or without other non-family persons, of two or more families sharing a dwelling, or a group of unrelated persons, or of one person living alone.

Marital Status: Refers to the conjugal status of a person: Married (including common-law); Separated; Divorced; Widowed; Never married (single).

Occupied private dwelling: Refers to a private dwelling in which a person or a group of persons are permanently residing. Also included are private dwellings whose usual residents are temporarily absent on Census Day.

Private dwelling: Refers to a separate set of living quarters with a private entrance either from outside or from a common hall, lobby, vestibule or stairway inside the building. The entrance to the dwelling must be one than can be used without passing through the living quarters of someone else.

Private household: Refers to a group of persons (other than foreign residents) who occupy a private dwelling and do not have a usual place of residence elsewhere in Canada.

Appendix B – WA and EA Level Constraints Applied to the 1996 Census Weights

(Note: The 1996 Census second run additional constraints are flagged with an “#”.)

Person WA Level Constraints

- Total persons
- Total persons aged ≥ 15

- Males
- Males aged ≥ 15

- Persons aged 0 to 4
- Persons aged 5 to 9
- Persons aged 10 to 14
- Persons aged 15 to 19
- Persons aged 20 to 24
- Persons aged 25 to 29
- Persons aged 30 to 34
- Persons aged 35 to 39
- Persons aged 40 to 44
- Persons aged 45 to 49
- Persons aged 50 to 54
- Persons aged 55 to 59
- Persons aged 60 to 64 #
- Persons aged 65 to 74
- Persons aged ≥ 75

- Married persons
- Single persons
- Divorced persons
- Widowed persons
- Common-law #

EA Level Constraints

- Total households in EA
- Total persons in EA

Household WA Level Constraints

- Households of size 1 #
- Households of size 2
- Households of size 3
- Households of size 4
- Households of size 5

Appendix C – Additional Information on Statistics Used in Sampling Bias Study

Let X represent the known value for a 2A characteristic at the census division (CD) level and let $\hat{X}^{(0)}$ represent the Horvitz-Thompson estimator of X . $\hat{X}^{(0)}$ was calculated by multiplying the unweighted sample total for the characteristic of each sampled EA by the inverse of the realized household sampling fraction for the EA, and then summing the results to the census division (CD) level. Non-sampled enumeration areas (EAs) were excluded from the analysis. The standard deviation of $\hat{X}^{(0)}$, $\text{std } \hat{X}^{(0)} = \sqrt{V(\hat{X}^{(0)})}$ was calculated under the assumption that simple random samples of households were drawn independently in each EA (in fact, independent systematic random samples were drawn). Consequently, the variances were calculated at the EA level and summed to the CD level. The population S^2 values were used in the variance calculations. See Cochran, 1977, pp. 23-24, for variance formulas for person and family characteristics, and pp. 50-52, for variance formulas for household and dwelling characteristics.

Since the $\hat{X}^{(0)}$ values are Horvitz-Thompson estimators, they are unbiased for X . Sampling was done independently in different EAs. Therefore, the $\hat{X}^{(0)}$ values are the sum of n independent random variables, where n is the number of EAs in the CD. Since 90% of the CDs had more than 25 EAs with an average of 140, n is quite large in most CDs. Thus, according to the central limit theorem, $Z^{(0)} = (\hat{X}^{(0)} - X) / \text{std } (\hat{X}^{(0)})$ should follow an approximately normal (0.1) distribution (see Kendall and Stuart, 1963, p. 193). This, however, would not be the case if 2B responses were significantly biased for any reason.

The $Z^{(0)}$ values were produced for all 281 sampled CDs in Canada, for the 2A characteristics given in Chapter V. In order to evaluate the normality of the $Z^{(0)}$ values at the CD level, histograms of the $Z^{(0)}$ values overlaid with a normal PDF (Probability Density Function) were produced.

In addition, to test whether $Z^{(0)}$ was being selected from a normal distribution whose mean is zero (i.e. the sample

selection procedure was unbiased), the mean $\bar{Z}^{(0)} = \sum_{i=1}^m Z_i^{(0)} / m$ was calculated where $m = 281$ (the number of CDs)

and $Z_i^{(0)}$ is the value of $Z^{(0)}$ for the i^{th} CD. In addition, the standard deviation of the $Z_i^{(0)}$ was determined where

$\text{std}^2(Z^{(0)}) = \sum_{i=1}^m (Z_i^{(0)} - \bar{Z}^{(0)})^2 / (m-1)$. Then the T statistic $T_z = \sqrt{m} \bar{Z}^{(0)} / \text{std}(Z^{(0)})$ was calculated. If the sample

selection procedure was unbiased, T should follow Student's t distribution with $m-1$ degrees of freedom. The probability of $|T_z| > 1.960$ if the sample selection procedure was unbiased is less than 0.05. Thus, if $|T_z| > 1.960$, the hypothesis that the sample selection procedure was unbiased will be rejected and the difference between the sample estimate and the population count will be said to be statistically significant at the 5% level.

Appendix D – 1986 Standard Error Adjustment Factors at National or Provincial Level and Percentiles of Weighting Area Level Factors

Characteristics	National or Provincial Factor	Percentiles of WA Level Factors					
		1	50	75	90	95	99
Population Characteristics							
Age							
Age groups 0-4, 5-9, 10-14, 15-19, 20-24, 25-29	0.18	0.05	0.19	0.29	0.35	0.49	0.52
Age groups 30-34, 35-44, 45-54, 55-59, 60-64, 5+, 15+	0.36	0.13	0.33	0.46	0.51	0.56	0.61
Age group 65+	0.00	-	-	-	-	-	-
Sex	0.00	-	-	-	-	-	-
Marital status							
Single, married (excluding separated)	0.25	0.04	0.23	0.31	0.42	0.49	0.55
Separated, divorced, widowed	0.88	0.55	0.84	0.98	1.06	1.15	1.20
Highest level of schooling							
Highest degree, certificate or diploma/total years of schooling	0.90	0.75	0.95	1.06	1.14	1.19	1.25
Major field of study	1.20	0.84	1.16	1.22	1.28	1.35	1.43
Mobility status							
Non-movers	1.21	0.83	1.23	1.27	1.32	1.36	1.41
Movers (migrants, non-migrants)	1.61	0.90	1.60	1.75	1.85	1.97	2.09
Period of immigration							
Before 1946, 1946-1966	0.98	0.76	1.02	1.10	1.22	1.37	1.45
1967-1977, 1978-1982, 1983-1986	1.51	0.80	1.45	1.55	1.78	1.90	2.11
Age at immigration	1.10	0.71	1.15	1.29	1.38	1.44	1.54

Characteristics	National or Provincial Factor	Percentiles of WA Level Factors					
		1	50	75	90	95	99
Population Characteristics							
Place of birth							
Born in Canada	1.09	0.82	1.08	1.16	1.18	1.20	1.21
Born outside Canada	1.35	1.11	1.34	1.43	1.60	1.67	1.75
Immigrant/Non-immigrant population	1.12	0.81	1.10	1.24	1.38	1.46	1.52
Citizenship							
Canada, by birth	1.13	0.88	1.14	1.17	1.20	1.27	1.32
Other	1.59	1.04	1.40	1.65	1.88	1.95	2.12
Ethnic origin							
English, French	1.20	0.73	1.16	1.25	1.31	1.40	1.46
Other	1.65	1.07	1.57	1.70	1.89	1.99	2.11
Home language							
English, French, English and French, English and non-official languages	1.12	0.50	1.09	1.35	1.75	1.89	2.09
Other language groups	1.76	0.99	1.68	1.89	2.01	2.20	2.41
Official language							
English, French, English and French	1.05	0.69	1.01	1.18	1.31	1.42	1.58
Other language groups	1.49	0.90	1.50	1.68	1.76	1.79	1.91
Mother tongue – English							
Newfoundland, Prince Edward Island, Nova Scotia, British Columbia	0.92	0.24	0.96	1.45	1.62	1.90	2.23
Quebec	1.15	0.18	1.10	1.51	1.76	1.81	1.99
Other provinces	0.45	0.12	0.48	0.71	0.96	1.12	1.38
Canada	0.53	-	-	-	-	-	-
Mother tongue – French							
Quebec	0.42	0.14	0.45	0.52	0.61	0.76	0.91
New Brunswick	0.75	0.19	0.79	0.98	1.24	1.60	1.84
Other provinces	1.04	0.09	1.12	1.49	1.71	1.89	2.06
Canada	0.77	-	-	-	-	-	-

Characteristics	National or Provincial Factor	Percentiles of WA Level Factors					
		1	50	75	90	95	99
Population Characteristics							
Mother tongue – Other language groups	1.70	0.73	1.63	2.11	2.44	2.51	2.60
Industry/Occupation	0.92	0.25	0.80	1.13	1.25	1.31	1.38
Work activity in 1985	0.89	0.62	0.92	1.14	1.22	1.29	1.31
Weeks worked in 1985	0.94	0.68	0.99	1.18	1.29	1.33	1.39
Hours worked in reference week	0.83	0.63	0.85	1.01	1.14	1.19	1.24
Year last worked							
In 1986, in 1985, before 1985	0.89	0.60	0.94	0.99	1.05	1.11	1.20
Never worked	1.18	0.80	1.15	1.34	1.43	1.50	1.67
Class of worker							
Paid workers	0.72	0.56	0.75	0.86	0.93	0.95	0.98
Self-employed, unincorporated, unpaid family workers	0.93	0.68	0.96	1.08	1.13	1.15	1.18
Labour force status participation							
Employed	0.75	0.59	0.76	0.83	0.86	0.91	0.93
Unemployed	1.06	0.76	1.04	1.14	1.20	1.27	1.38
Not in labour force	1.25	0.91	1.30	1.43	1.50	1.58	1.63
Major source of income							
Wages and salaries	0.65	0.42	0.67	0.80	0.85	0.87	0.92
Other	1.05	0.71	1.00	1.12	1.17	1.20	1.24
Major source of income							
Wages and salaries	0.65	0.42	0.67	0.80	0.85	0.87	0.92
Other	1.05	0.71	1.00	1.12	1.17	1.20	1.24

Characteristics	National or Provincial Factor	Percentiles of WA Level Factors					
		1	50	75	90	95	99
Population Characteristics							
Disability							
Limited at home, school and work	0.94	0.69	0.96	1.11	1.29	1.34	1.42
Not limited	0.61	0.41	0.58	0.69	0.74	0.78	0.81
Census family status							
Husband, wife, child	0.20	0.05	0.20	0.24	0.26	0.28	0.31
Female lone parent	0.45	0.14	0.43	0.51	0.55	0.61	0.68
Male lone parent, non-member of a census family	0.68	0.35	0.65	0.79	0.89	0.99	1.14
Economic family status							
Husband, wife	0.14	0.06	0.16	0.21	0.28	0.34	0.36
Lone parent, child	0.32	0.16	0.34	0.39	0.44	0.47	0.53
Other family members	0.74	0.24	0.70	0.84	1.03	1.09	1.18
Number of persons in the census family	0.04	0.00	0.00	0.05	0.07	0.09	0.11
Number of persons in the economic family	0.18	0.08	0.19	0.24	0.33	0.41	0.45
Age of husband, wife or reference person of economic family	1.42	0.80	1.37	1.53	1.60	1.78	1.91
All other population characteristics	1.00	-	-	-	-	-	-
Household and Dwelling Characteristics							
Structural type							
Single detached	0.33	0.05	0.35	0.55	0.67	0.75	0.89
Apartment less than 5 storeys	0.57	0.12	0.56	0.70	0.83	0.99	1.26
Other	0.91	0.18	0.88	0.99	1.18	1.23	1.32

Characteristics	National or Provincial Factor	Percentiles of WA Level Factors					
		1	50	75	90	95	99
Household and Dwelling Characteristics							
Tenure	0.00	-	-	-	-	-	-
Period of construction	0.78	0.61	0.75	0.82	0.89	0.99	1.24
Main type of heating equipment/principal heating fuel	0.87	0.18	0.86	1.04	1.12	1.25	1.32
Central heating equipment							
With	0.42	0.09	0.38	0.54	0.60	0.70	0.89
Without	0.78	0.23	0.79	0.91	1.03	1.12	1.20
Household size							
One-person household	0.00	-	-	-	-	-	-
Other	0.76	0.19	0.72	1.09	1.17	1.21	1.30
Number of rooms	0.80	0.57	0.78	0.90	0.97	1.10	1.20
Age of household maintainer							
25-34, 55-64, 65-74, 75+	0.25	0.06	0.24	0.35	0.48	0.53	0.62
0-24, 35-44, 45-54	0.92	0.38	0.90	1.05	1.14	1.21	1.30
Sex of household maintainer							
Male	0.20	0.09	0.24	0.31	0.34	0.36	0.37
Female	0.47	0.16	0.43	0.54	0.64	0.74	0.89
Gross rent/gross rent as a percentage of household income	0.75	0.48	0.79	0.91	0.94	0.96	1.01
Owner's major payments/owner's major payments as a percentage of household income	0.84	0.62	0.87	0.95	1.01	1.04	1.11

Characteristics	National or Provincial Factor	Percentiles of WA Level Factors					
		1	50	75	90	95	99
Household and Dwelling Characteristics							
Household income	0.75	0.51	0.73	0.82	0.90	0.95	1.03
Value of dwelling	0.90	0.67	0.91	1.00	1.05	1.12	1.18
Registered condominium							
Part	0.63	0.18	0.59	0.84	0.93	1.11	1.30
Not part	0.15	0.07	0.14	0.19	0.28	0.39	0.47
Household type – One-family households							
Without additional persons	0.22	0.05	0.20	0.27	0.33	0.36	0.40
With additional persons	0.50	0.20	0.48	0.61	0.72	0.74	0.79
Household type – Non-family households	0.00	-	-	-	-	-	-
Household type – Other	1.12	0.54	1.05	1.26	1.40	1.51	1.67
All other household and dwelling characteristics	1.00	-	-	-	-	-	-
Census Family Characteristics							
Census family structure							
Husband and wife	0.20	0.09	0.21	0.26	0.29	0.33	0.36
Male lone parent	0.64	0.21	0.62	0.81	0.84	0.91	1.04
Female lone parent	0.46	0.19	0.45	0.57	0.65	0.69	0.74
Census family type							
Primary family	0.23	0.04	0.24	0.28	0.31	0.34	0.39
Secondary family	0.90	0.62	0.93	1.15	1.28	1.33	1.40
Age groups of children at home	0.78	0.40	0.70	0.91	0.98	1.09	1.19

Characteristics	National or Provincial Factor	Percentiles of WA Level Factors					
		1	50	75	90	95	99
Census Family Characteristics							
Labour force activity of husband, wife or lone parent							
Husband, lone parent, husband and wife in labour force	0.40	0.23	0.43	0.50	0.55	0.59	0.71
Wife in labour force	0.61	0.41	0.60	0.68	0.74	0.78	0.82
Other	0.72	0.30	0.68	0.80	0.90	0.99	1.12
Work activity in 1985 of husband, wife or lone parent							
Worked in 1985	0.48	0.11	0.45	0.50	0.54	0.57	0.59
Did not work in 1985	0.93	0.60	0.90	1.04	1.18	1.26	1.30
All other census family characteristics	1.00	-	-	-	-	-	-
Economic Family Characteristics							
Economic family structure							
Husband and wife families	0.29	0.13	0.30	0.36	0.48	0.56	0.68
Non-husband and wife families	0.56	0.35	0.50	0.66	0.81	0.90	1.06
Mother tongue of family reference person – English							
Newfoundland, Prince Edward Island, British Columbia	0.25	0.09	0.20	0.31	0.45	0.66	0.91
Quebec	0.49	0.25	0.47	0.50	0.69	0.83	1.05
Other provinces	0.18	0.07	0.19	0.22	0.24	0.27	0.31
Canada	0.27	-	-	-	-	-	-
Mother tongue of family reference person – English							
Newfoundland, Prince Edward Island, British Columbia	0.25	0.09	0.20	0.31	0.45	0.66	0.91
Quebec	0.49	0.25	0.47	0.50	0.69	0.83	1.05
Other provinces	0.18	0.07	0.19	0.22	0.24	0.27	0.31
Canada	0.27	-	-	-	-	-	-

Characteristics	National or Provincial Factor	Percentiles of WA Level Factors					
		1	50	75	90	95	99
Economic Family Characteristics							
Mother tongue of family reference person – Other than English or French							
Newfoundland, Nova Scotia	0.75	0.38	0.74	0.80	0.91	0.99	1.10
Other provinces	0.50	0.21	0.45	0.57	0.82	0.84	0.99
Canada	0.56	-	-	-	-	-	-
All other economic family characteristics	1.00	-	-	-	-	-	-

Appendix E – Products and Services

Packaging census data so they are meaningful and accessible to clients, whether they are government decision-makers, policy analysts, librarians, marketing specialists, researchers, students, etc., is the key to ensuring that the value of the data is maximized. There are several new product and service features for 1996.

1. Increased Accessibility Through Electronic Media

More clients asked for census materials to be available in electronic formats which can be used with personal computers. While some key printed products have been retained, more census data were produced on CD-ROM and on diskette. These formats contained Windows-based presentation and tabulation softwares to make the data easy to use. For the first time, clients were able to obtain information free of charge on the Internet through the Statistics Canada's Web site: <http://www.statcan.ca>.

2. Small Area Data Available Sooner

Census data at smaller levels of geography were made available much sooner than in previous years. On each release day, profile data were available for areas at the community levels (census subdivisions and census divisions) and, one month after their release, data for areas as small as census tracts, enumeration areas and forward sortation areas.

3. Census Tabulations Available by Postal Code

As part of the standard product line, basic summary tabulations and area profiles were available for forward sortation areas, which represent the first three characters of the postal code. Data for the full postal code can be obtained as a custom service, subject to confidentiality restrictions.

4. New Information Collected in 1996

For the first time, data will be published for unpaid household activities, place of work for all levels of geography, mode of transportation to work and population groups.

5. Improvement of Geography Products

Not only has the quality of many of the maps used for the release of census data been improved, a map series on federal electoral districts has been reintroduced. *GeoSuite* (formerly *GeoRef*), the Windows-based electronic tool which allows clients to explore the links between different levels of geography, has also been improved with the addition of enumeration area reference lists.

Bibliography

Census Operations Division produced the following portions of this report: Introduction, Appendices A and E, and Regional Reference Centres.

Béland, Y. "Results and Methodology of the 1986 Sampling Variance Study". Statistics Canada Internal Report, 1990.

Bankier, M., S. Rathwell, and M. Majkowski. "Two-step Generalized Least Squares Estimation in the 1991 Canadian Census". Methodology Branch Working Paper, Social Survey Methods Division, 1992. Catalogue number 92-007E.

Bankier, M., A.-M. Houle, and M. Luc. Calibration Estimation in the 1991 and 1996 Canadian Censuses. Proceedings of the Survey Methods Research Section, American Statistical Association, pp. 66-75. 1997.

Brackstone, G.J. and J.N.K. Rao. "An Investigation of Raking Ratio Estimators". Sankhya, Volume 41, Series C, Pt. 2, p. 97-114. 1979.

Cochran, W. Sampling Techniques 3rd Edition. John Wiley and Sons: Toronto, 1977.

Dominion Bureau of Statistics. Eighth Census of Canada, 1941, Administrative Report of the Dominion Statistician, Ottawa: King's Printer, 1945.

Dominion Bureau of Statistics. Ninth Census of Canada, 1951, Vol. XI, Administrative Report, Ottawa: Queen's Printer, 1955.

Dominion Bureau of Statistics. Sampling in the Census. S.M.S.03.5, 1968.

Dominion Bureau of Statistics. 1961 Census of Canada. General Review, Bulletin 7.2-12, Ottawa: Queen's Printer, 1970. Catalogue number 99-537.

Fellegi, I.P. "Response Variance and its Estimation". Journal of the American Statistical Association, 59, pp. 1016-1041. 1964.

Hansen, M.H., W.N. Hurwitz, and M.A. Bershad. "Measurement Errors in Censuses and Surveys". Bulletin of the International Statistical Institute, 38, pp. 359-374. 1959.

Kendall, M.G. and A. Stuart. "The Advanced Theory of Statistics", Volume 1, Charles Griffin and Company Limited, London, 1963.

- Kruszynski, G. Evaluation of the 1996 Weighting Factor. Internal Report, Geography Division, Statistics Canada, 1999.
- Majkowski, M. "1991 Census 2A/2B Discrepancies". Statistics Canada Internal Report, 1992a.
- Majkowski, M. "Investigation into Large Population/Estimate Differences in the 1991 Census". Statistics Canada Report, 1994.
- Royce, D. "The Use of Sampling in the 1981 Canadian Census". Statistics Canada Internal Report, 1983.
- Sarndal, C., B. Swensson, and J. Wretman. Model Assisted Survey Sampling. Springer-Verlag: New York, 1992.
- Statistics Canada. 1971 Census of Canada. General Review, Vol. VI, Part 1, Ottawa, 1976. Catalogue number 99-740.
- Statistics Canada. 1976 Census of Canada. Administrative Report, Part 1, Ottawa, 1980. Catalogue number 99-850.
- Statistics Canada. 1976 Census of Canada. Quality of Data, Series 1: Sources of Error - Sampling and Weighting. Ottawa, 1980. Catalogue number 99-844.
- Statistics Canada. 1981 Census of Canada. Summary Guide: Total Population. Ottawa, 1983. Catalogue number 99-902.
- Statistics Canada. 1986 Census of Canada. Census Handbook. Ottawa, 1988. Catalogue number 99-104E.
- Statistics Canada. User's Guide to the Quality of 1986 Census Data: Sampling and Weighting. Ottawa, 1990. Catalogue number 99-136E.
- Statistics Canada. 1991 Census of Canada. Census Handbook. Ottawa, 1992. Catalogue number 92-305E.
- Thivierge, S. Bias in the 1996 Census Sample. Internal Report, Social Survey Methods Division, Statistics Canada, 1999.

Regional Reference Centres

Statistics Canada regional reference centres are located across the country. Each centre has a complete collection of current publications and reference documents which can be consulted or purchased, along with microcomputer diskettes, CD-ROMs, maps and other products and services, including CANSIM.

Each Reference Centre provides a wide range of additional services. On the one hand, the Dissemination Services: a free telephone enquiries line for the most recent basic data. On the other hand, Advisory Services: identification of your needs, establishing sources or availability of data, consolidation and integration of data coming from different sources and development of profiles, analysis of highlights or tendencies and, finally, training on products, services, Statistics Canada concepts and the use of statistical data.

For more information about the services provided by regional reference centres, you can call or visit the closest centre. The locations are listed below. If you are outside the local calling area, please dial the toll-free number.

National toll-free **enquiries** line (Canada and United States): 1 800 263-1136
Telecommunications device for the **hearing-impaired**: 1 800 363-7629
Toll-free **order only** line (Canada and United States): 1 800 267-6677
National toll-free **fax order** line (Canada and United States): 1 800 287-4369

List of Statistics Canada Regional Reference Centres

Atlantic Region

Serving the provinces of Newfoundland and Labrador, Nova Scotia, Prince Edward Island and New Brunswick

Statistics Canada

Advisory Services

1741 Brunswick Street

2nd floor, Box 11

Halifax, Nova Scotia B3J 3X8

Toll-free number: 1 800 263-1136

Local calls: (902) 426-5331

Fax number: (902) 426-9538

E-mail: atlantic.info@statcan.ca

Quebec Region

Serving the province of Quebec and the territory of Nunavut, except the National Capital Region

Statistics Canada

Advisory Services

200 René Lévesque Blvd W.

Guy Favreau Complex

4th floor, East Tower

Montréal, Quebec H2Z 1X4

Toll-free number: 1 800 263-1136

Local calls: (514) 283-5725

Fax number: (514) 283-9350

E-mail: louise.bernot@statcan.ca

National Capital Region

Serving the National Capital Region

Statistics Canada

Statistical Reference Centre

R.H. Coats Building Lobby

Holland Avenue

Ottawa, Ontario K1A 0T6

Toll-free number: 1 800 263-1136

Local calls: (613) 951-8116

Fax number: (613) 951-0581

E-mail: infostats@statcan.ca

If you live outside the local calling area, please dial the toll-free number for your region.

Ontario Region

Serving the province of Ontario, except the National Capital Region

Statistics Canada

Advisory Services

Arthur Meighen Building

10th floor

25 St. Clair Avenue East

Toronto, Ontario M4T 1M4

Toll-free number: 1 800 263-1136

Local calls: (416) 973-6586

Fax number: (416) 973-7475

Prairie Region

Serving the provinces of Manitoba, Saskatchewan, Alberta and the Northwest Territories

Serving the province of Manitoba:

Statistics Canada

Advisory Services

Via Rail Building, Suite 200

123 Main Street

Winnipeg, Manitoba R3C 4V9

Toll-free number: 1 800 263-1136

Local calls: (204) 983-4020

Fax number: (204) 983-7543

E-mail: statswpg@solutions.net

Serving the province of Saskatchewan:

Statistics Canada

Advisory Services

Park Plaza, Suite 440

2365 Albert Street

Regina, Saskatchewan S4P 4K1

Toll-free number: 1 800 263-1136

Local calls: (306) 780-5405

Fax number: (306) 780-5403

E-mail: statcan@sk.sympatico.ca

Serving Northern Alberta and the Northwest Territories:

Statistics Canada
Advisory Services
Park Square, 15th floor
10001 Bellamy Hill
Edmonton, Alberta T5J 3B6
Toll-free number: 1 800 263-1136
Local calls: (780) 495-3027
Fax number: (780) 495-5318
E-mail: ewieall@statcan.ca

Serving Southern Alberta:

Statistics Canada
Advisory Services
Discovery Place, Room 201
3553-31 Street N.W.
Calgary, Alberta T2L 2K7
Toll-free number: 1 800 263-1136
Local calls: (403) 292-6717
Fax number: (403) 292-4958
E-mail: degagnej@cadvision.com

Pacific Region

Serving the province of British Columbia and the Yukon Territory

Statistics Canada
Advisory Services
Library Square Office Tower
600 - 300 West Georgia Street
Vancouver, British Columbia V6B 6C7
Toll-free number: 1 800 263-1136
Local calls: (604) 666-3691
Fax number: (604) 666-4863
E-mail: stcvan@statcan.ca

Appels sans frais : 1 800 263-1136
Appels locaux : (306) 780-5405
Télécopieur : (306) 780-5403
Courrier électronique : statcan@sk.sympatico.ca

Pour le nord de l'Alberta et les Territoires du Nord-Ouest :
Statistique Canada
Services-conseils
Park Square, 15^e étage
10001, Bellamy Hill
Edmonton (Alberta) T5J 3B6
Appels sans frais : 1 800 263-1136
Appels locaux : (780) 495-3027
Télécopieur : (780) 495-5318
Courrier électronique : ewieall@statcan.ca

Pour le sud de l'Alberta :
Statistique Canada
Services-conseils
Discovery Place, pièce 201
3553, 31^e rue N.-O.
Calgary (Alberta) T2L 2K7
Appels sans frais : 1 800 263-1136
Appels locaux : (403) 292-6717
Télécopieur : (403) 292-4958
Courrier électronique : degagnej@cadvision.com

Région du Pacifique
Dessert la Colombie-Britannique et le Territoire du Yukon
Statistique Canada
Services-conseils
Library Square Office Tower
600 - 300, rue Georgia Ouest
Vancouver (Colombie-Britannique) V6B 6C7
Appels sans frais : 1 800 263-1136
Appels locaux : (604) 666-3691
Télécopieur : (604) 666-4863
Courrier électronique : stcvan@statcan.ca

Appels sans frais : 1 800 263-1136
Appels locaux : (514) 283-5725
Télécopieur : (514) 283-9350
Courrier électronique : louise.bourmot@statcan.ca

Région de la Capitale nationale

Dessert la région de la Capitale nationale
Statistique Canada
Centre de consultation statistique
Immeuble R.-H.-Coats, Rez-de-chaussée
Avenue Holland

Ottawa (Ontario) K1A 0T6
Appels sans frais : 1 800 263-1136
Appels locaux : (613) 951-8116

Télécopieur : (613) 951-0581
Courrier électronique : infostats@statcan.ca

Si vous demeurez à l'extérieur de la zone de communication locale, composez le numéro sans frais de votre région.

Région de l'Ontario

Dessert tout l'Ontario, sauf la région de la Capitale nationale
Statistique Canada
Services-conseils

Immeuble Arthur-Meighen, 10^e étage
25, avenue St. Clair Est

Toronto (Ontario) M4T 1M4
Appels sans frais : 1 800 263-1136
Appels locaux : (416) 973-6586

Télécopieur : (416) 973-7475

Région des Prairies

Dessert le Manitoba, la Saskatchewan, l'Alberta et les Territoires du Nord-Ouest

Pour le Manitoba

Statistique Canada
Services-conseils

Immeuble Via Rail, pièce 200
123, rue Main

Winnipeg (Manitoba) R3C 4V9
Appels sans frais : 1 800 263-1136
Appels locaux : (204) 983-4020

Télécopieur : (204) 983-7543

Courrier électronique : statswpg@solutions.net

Pour la Saskatchewan :

Statistique Canada
Services-conseils

Park Plaza, suite 440
2365, rue Albert

Regina (Saskatchewan) S4P 4K1

Centres régionaux de consultation

Les centres régionaux de consultation de Statistique Canada sont situés un peu partout au pays et chacun possède une série complète des publication courantes et des documents de référence que les clients peuvent consulter sans frais ou acheter. On y trouve aussi des disquettes pour micro-ordinateurs, des CD-ROM, des cartes et d'autres produits et services, dont CANSIM.

Chaque centre de consultation vous offre une gamme de produits et services additionnels. D'une part, les Services de diffusion : service téléphonique gratuit de renseignements pour les données de base les plus récentes. D'autre part, les conseils : identification de vos besoins, détermination des sources ou de la disponibilité des données, consolidation et intégration de données provenant de diverses sources et élaboration de profils, analyse de faits saillants ou de tendances et, pour terminer, formation sur les produits, services et concepts de Statistique Canada ainsi que l'utilisation de données statistiques.

Pour en savoir davantage sur les services offerts dans les centres régionaux de consultation, il suffit de communiquer avec le centre de sa région, dont la liste se trouve ci-dessous. Les personnes qui habitent à l'extérieur de la zone de communication locale disposent d'un numéro sans frais.

Numéro sans frais pour les **demandes de renseignements** (Canada et États-Unis) : 1 800 263-1136
Appareils de télécommunications pour les **malentendants** : 1 800 363-7629
Numéro sans frais pour **commander seulement** (Canada et États-Unis) : 1 800 267-6677
Numéro sans frais de commande par **télécopieur** (Canada et États-Unis) : 1 800 287-4369

Liste des centres régionaux de consultation de Statistique Canada

Région de l'Atlantique
Dessert Terre-Neuve et le Labrador, la Nouvelle-Écosse, l'Île-du-Prince-Édouard et le Nouveau-Brunswick
Statistique Canada
Services-conseils
1741, rue Brunswick
2^e étage, boîte 11
Halifax (Nouvelle-Écosse) B3J 3X8
Appels sans frais : 1 800 263-1136
Appels locaux : (902) 426-5331
Télécopieur : (902) 426-9538
Courrier électronique : atlantic.info@statcan.ca

Région du Québec
Dessert tout le Québec, sauf la région de la Capitale nationale et le territoire du Nunavut
Statistique Canada
Services-conseils
200, boul. René-Lévesque Ouest
Complexe Guy-Favreau
4^e étage, Tour Est
Montréal (Québec) H2Z 1X4

Kendall, M.G. et A. Stuart. «The Advanced Theory of Statistics», volume 1, Charles Griffin and Company Limited, London, 1963.

Kruszynski, G. Evaluation of the 1996 Weighting Factor. Rapport interne de la Division de la géographie, Statistique Canada, 1999.

Majkowski, M. «1991 Census 2A/2B Discrepancies». Rapport interne de Statistique Canada, 1992a.

Majkowski, M. «Investigation into Large Population/Estimate Differences in the 1991 Census». Rapport de Statistique Canada, 1994.

Royce, D. «The Use of Sampling in the 1981 Canadian Census». Rapport interne de Statistique Canada, 1983.

Sandal, C., B. Swensson, et J. Wretman. «Model Assisted Survey Sampling». Springer-Verlag : New York, 1992.

Statistique Canada. Recensement du Canada de 1971. Revue générale, vol. VI, partie 1, numéro 99-740 au catalogue, Ottawa, 1976.

Statistique Canada. Recensement du Canada de 1976. Rapport administratif, partie 1, numéro 99-850 au catalogue, Ottawa, 1980.

Statistique Canada. Recensement du Canada de 1976. «Quality of Data, Series 1 : Sources of Error - Sampling and Weighting». Numéro 99-844 au catalogue, Ottawa, 1980.

Statistique Canada. Recensement du Canada de 1981. Guide sommaire : Population totale. Numéro 99-902 au catalogue, Ottawa, 1983.

Statistique Canada. Recensement du Canada de 1986. Le Recensement en bref. Numéro 99-104F au catalogue, Ottawa, 1988.

Statistique Canada. Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération. Numéro 99-136F au catalogue, Ottawa, 1990.

Statistique Canada. Recensement du Canada de 1991. Le recensement de 1991 en bref, numéro 92-305F au catalogue, Ottawa, 1992.

Thievery, S. «Bias in the 1996 Census Sample». Rapport interne de la Division des méthodes d'enquêtes sociales, Statistique Canada, 1999.

Centres régionaux de consultation

Les centres régionaux de consultation de Statistique Canada sont situés un peu partout au pays et chacun possède une série complète des publications courantes et des documents de référence que les clients peuvent consulter sans frais ou acheter. On y trouve aussi des disquettes pour micro-ordinateurs, des CD-ROM, des cartes et d'autres produits et services, dont CANSIM.

Chaque centre de consultation vous offre une gamme de produits et services additionnels. D'une part, les Services de diffusion : service téléphonique gratuit de renseignements pour les données de base les plus récentes. D'autre part, les conseils : identification de vos besoins, détermination des sources ou de la disponibilité des données, consolidation et intégration de données provenant de diverses sources et élaboration de profils, analyse de faits saillants ou de tendances et, pour terminer, formation sur les produits, services et concepts de Statistique Canada ainsi que l'utilisation de données statistiques.

Pour en savoir davantage sur les services offerts dans les centres régionaux de consultation, il suffit de communiquer avec le centre de sa région, dont la liste se trouve ci-dessous. Les personnes qui habitent à l'extérieur de la zone de communication locale disposent d'un numéro sans frais.

Numéro sans frais pour les **demandes de renseignements** (Canada et États-Unis) : 1 800 263-1136
Appareils de télécommunications pour les **malentendants** : 1 800 363-7629
Numéro sans frais pour **commander seulement** (Canada et États-Unis) : 1 800 267-6677
Numéro sans frais de commande par **télécopieur** (Canada et États-Unis) : 1 800 287-4369

Liste des centres régionaux de consultation de Statistique Canada

Région de l'Atlantique
Dessert Terre-Neuve et le Labrador, la Nouvelle-Écosse, l'Île-du-Prince-Édouard et le Nouveau-Brunswick
Statistique Canada
Services-conseils
1741, rue Brunswick
2^e étage, boîte 11
Halifax (Nouvelle-Écosse) B3J 3X8
Appels sans frais : 1 800 263-1136
Appels locaux : (902) 426-5331
Télécopieur : (902) 426-9538
Courrier électronique : atlantic.info@statcan.ca

Région du Québec
Dessert tout le Québec, sauf la région de la Capitale nationale et le territoire du Nunavut
Statistique Canada
Services-conseils
200, boul. René-Lévesque Ouest
Complexe Guy-Favreau
4^e étage, Tour Est
Montréal (Québec) H2Z 1X4

Kendall, M.G. et A. Stuart. «The Advanced Theory of Statistics», volume 1, Charles Griffin and Company Limited, London, 1963.

Kruszynski, G. Evaluation of the 1996 Weighting Factor. Rapport interne de la Division de la géographie, Statistique Canada, 1999.

Majkowski, M. «1991 Census 2A/2B Discrepancies». Rapport interne de Statistique Canada, 1992a.

Majkowski, M. «Investigation into Large Population/Estimate Differences in the 1991 Census». Rapport de Statistique Canada, 1994.

Royce, D. «The Use of Sampling in the 1981 Canadian Census». Rapport interne de Statistique Canada, 1983.

Sarnadal, C., B. Swensson, et J. Wretman. «Model Assisted Survey Sampling». Springer-Verlag : New York, 1992.

Statistique Canada. Recensement du Canada de 1971. Revue générale, vol. VI, partie 1, numéro 99-740 au catalogue, Ottawa, 1976.

Statistique Canada. Recensement du Canada de 1976. Rapport administratif, partie 1, numéro 99-850 au catalogue, Ottawa, 1980.

Statistique Canada. Recensement du Canada de 1976. «Quality of Data, Series 1 : Sources of Error - Sampling and Weighting». Numéro 99-844 au catalogue, Ottawa, 1980.

Statistique Canada. Recensement du Canada de 1981. Guide sommaire : Population totale. Numéro 99-902 au catalogue, Ottawa, 1983.

Statistique Canada. Recensement du Canada de 1986. Le Recensement en bref. Numéro 99-104F au catalogue, Ottawa, 1988.

Statistique Canada. Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération. Numéro 99-136F au catalogue, Ottawa, 1990.

Statistique Canada. Recensement du Canada de 1991. Le recensement de 1991 en bref, numéro 92-305F au catalogue, Ottawa, 1992.

Thievery, S. «Bias in the 1996 Census Sample». Rapport interne de la Division des méthodes d'enquêtes sociales, Statistique Canada, 1999.

Bibliographie

La Division des opérations du recensement a produit les parties suivantes du présent rapport : l'Introduction, les annexes A et E ainsi que les Centres régionaux de consultation.

Béland, Y. «Results and Methodology of the 1986 Sampling Variance Study». Rapport interne de Statistique Canada, 1990.

Bankier, M., S. Rathwell, et M. Majkowski. «Two-step Generalized Least Squares Estimation in the 1991 Canadian Census». Document de travail de la Direction de la méthodologie, Division des méthodes d'enquêtes sociales, 1992. Numéro 92-007E au catalogue.

Bankier, M., A.-M. Houle, et M. Luc. Calibration Estimation in the 1991 and 1996 Canadian Censuses. Proceedings of the Survey Methods Research Section, American Statistical Association (1997), pages 66-75.

Brackstone, G.J. et J.N.K. Rao. «An Investigation of Raking Ratio Estimators». Sankhya, volume 41, série C, partie 2, pages 97-114, 1979.

Bureau fédéral de la statistique. Huitième recensement du Canada, 1941, Rapport administratif du statisticien en chef, Ottawa : King's Printer, 1945.

Bureau fédéral de la statistique. Neuvième recensement du Canada, 1951, vol. XI, Rapport administratif, Ottawa : Queen's Printer, 1955.

Bureau fédéral de la statistique. Sampling in the Census. S.M.S.03.5, 1968.

Bureau fédéral de la statistique. Recensement du Canada de 1961. Revue générale, Bulletin 7.2-12, numéro 99-537 au catalogue, Ottawa : Queen's Printer, 1970.

Cochran, W. «Sampling Techniques 3rd Edition». John Wiley and Sons : Toronto, 1977.

Fellegi, I.P. «Response Variance and its Estimation». Journal of the American Statistical Association, 59, pages 1016-1041, 1964.

Hansen, M.H., W.N. Hurwitz, et M.A. Bershad. «Measurement Errors in Censuses and Surveys». Bulletin of the International Statistical Institute, 38, pages 359-374, 1959.

Annexe E – Produits et services

Pour mettre en valeur les données du recensement, on doit les présenter de façon qu'elles soient pertinentes et accessibles pour les clients, qu'ils soient décideurs, analystes de politiques, bibliothécaires, spécialistes en commercialisation, chercheurs, étudiants, etc. Plusieurs nouveaux produits et services sont offerts à l'occasion du recensement de 1996.

1. Accessibilité accrue sur support électronique

Un nombre accru de clients ont demandé que les produits du recensement soient offerts sur support électronique utilisable sur des ordinateurs personnels. Bien qu'on ait conservé certains produits clés sur support imprimé, un plus grand nombre de données du recensement sont offertes sur CD-ROM et sur disquettes. Ces supports contiennent des logiciels de présentation et de totalisation pour environnement Windows, qui rendent les données faciles à utiliser. Pour la première fois, les clients ont pu obtenir des données gratuitement dans Internet, au site Web de Statistique Canada : <http://www.statcan.ca>.

2. Données régionales diffusées plus tôt

Les données du recensement à l'échelon des petites régions géographiques ont été diffusées beaucoup plus tôt que dans les années antérieures. Le jour même de chaque diffusion, les données des profils étaient accessibles à l'échelon des collectivités (subdivisions et divisions de recensement) et, un mois plus tard, les données pour des régions aussi petites que les secteurs de recensement, les secteurs de dénombrement et les régions de tri d'acheminement étaient diffusées.

3. Totalisations du recensement accessibles selon le code postal

Comme partie intégrante des produits normalisés, les tableaux sommaires de base et les profils de secteurs ont été diffusés pour les régions de tri d'acheminement, qui correspondent aux trois premiers caractères du code postal. Les données selon le code postal complet peuvent être obtenues par l'intermédiaire d'une demande de produit personnalisé, sous réserve des restrictions en matière de confidentialité.

4. Nouvelles données recueillies en 1996

Pour la première fois, des données sont diffusées sur les activités à la maison non rémunérées, le lieu de travail pour tous les échelons géographiques, le mode de transport utilisé pour se rendre au travail et les groupes de population.

5. Produits géographiques améliorés

On a non seulement amélioré la qualité de nombreuses cartes servant à la diffusion des données du recensement, mais aussi repris la production d'une série de cartes des circonscriptions électorales fédérales. Le produit *GéoSuite* (auparavant GéoRét), un outil électronique pour environnement Windows qui permet aux utilisateurs d'explorer les liens entre différents échelons géographiques, a également été amélioré par l'ajout de listes de référence des secteurs de dénombrement.

Caractéristiques	Facteur national ou provincial	Centiles des facteurs au niveau de la RP					
		1	50	75	90	95	99
Caractéristiques de la famille économique							
Langue maternelle de la famille de la personne de référence – Autre que l'anglais et le français							
Terre-Neuve, Nouvelle-Écosse	0,75	0,38	0,74	0,80	0,91	0,99	1,10
Autres provinces	0,50	0,21	0,45	0,57	0,82	0,84	0,99
Canada	0,56	-	-	-	-	-	-
Toutes les autres caractéristiques de la famille économique	1,00	-	-	-	-	-	-

Caractéristiques	Facteur	Centiles des facteurs au niveau de la RP					
		national ou provincial	1	50	75	90	95
Caractéristiques de la famille de recensement							
Travail en 1985 du conjoint, de la conjointe ou du parent seul	0,48	0,11	0,45	0,50	0,54	0,57	0,59
A travaillé en 1985							
N'a pas travaillé en 1985	0,93	0,60	0,90	1,04	1,18	1,26	1,30
Toutes les autres caractéristiques de la famille de recensement							
Toutes les autres caractéristiques de la famille de recensement	1,00	-	-	-	-	-	-
Caractéristiques de la famille économique							
Structure de la famille économique	0,29	0,13	0,30	0,36	0,48	0,56	0,68
Familles conjoint-conjointe							
N'est pas une famille conjoint-conjointe	0,56	0,35	0,50	0,66	0,81	0,90	1,06
Langue maternelle de la famille de la personne de référence – Anglais							
Terre-Neuve, Île-du-Prince-Édouard, Colombie-Britannique	0,25	0,09	0,20	0,31	0,45	0,66	0,91
Québec	0,49	0,25	0,47	0,50	0,69	0,83	1,05
Autres provinces	0,18	0,07	0,19	0,22	0,24	0,27	0,31
Canada	0,27	-	-	-	-	-	-
Langue maternelle de la famille de la personne de référence – Français							
Québec	0,12	0,04	0,13	0,17	0,21	0,29	0,36
Autres provinces	0,88	0,30	0,90	1,07	1,21	1,28	1,35
Canada	0,40	-	-	-	-	-	-

Caractéristiques	Facteur	Centiles des facteurs au niveau de la RP					
		national ou provincial	1	50	75	90	95
logements							
Caractéristiques des ménages et des							
Type de logement – Ménages non familiaux	0,00	-	-	-	-	-	-
Type de logement – Autres	1,12	0,54	1,05	1,26	1,40	1,51	1,67
Toutes les autres caractéristiques des ménages et des logements	1,00	-	-	-	-	-	-
Caractéristiques de la famille de recensement							
Structure de la famille de recensement							
Conjoint et conjointe	0,20	0,09	0,21	0,26	0,29	0,33	0,36
Parent seul de sexe masculin	0,64	0,21	0,62	0,81	0,84	0,91	1,04
Parent seul de sexe féminin	0,46	0,19	0,45	0,57	0,65	0,69	0,74
Genre de famille de recensement							
Famille principale	0,23	0,04	0,24	0,28	0,31	0,34	0,39
Famille secondaire	0,90	0,62	0,93	1,15	1,28	1,33	1,40
Groupes d'âge des enfants à la maison							
0,78	0,40	0,40	0,70	0,91	0,98	1,09	1,19
Activité du conjoint, de la conjointe ou du parent seul							
Conjoint, parent seul, conjoint et conjointe dans la population active	0,61	0,41	0,60	0,68	0,74	0,78	0,82
Autres	0,72	0,30	0,68	0,80	0,90	0,99	1,12

Caractéristiques	Facteur	Centiles des facteurs au niveau de la RP					logements	
		national ou provincial	1	50	75	90		95
Caractéristiques des ménages et des logements	Nombre de pièces	0,80	0,57	0,78	0,90	0,97	1,10	1,20
	Âge du soutien du ménage	0,25	0,06	0,24	0,35	0,48	0,53	0,62
	0 à 24, 35 à 44, 45 à 54	0,92	0,38	0,90	1,05	1,14	1,21	1,30
	25 à 34, 55 à 64, 65 à 74, 75+							
	Sexe du soutien du ménage	0,20	0,09	0,24	0,31	0,34	0,36	0,37
	Hommes							
	Femmes	0,47	0,16	0,43	0,54	0,64	0,74	0,89
	Loyer brut/Loyer brut, sous forme de pourcentage du revenu du ménage	0,75	0,48	0,79	0,91	0,94	0,96	1,01
	Principales dépenses de propriété/	0,84	0,62	0,87	0,95	1,01	1,04	1,11
	Principales dépenses de propriété, sous forme de pourcentage du revenu du ménage							
Type de logement – Ménages	Revenu du ménage	0,75	0,51	0,73	0,82	0,90	0,95	1,03
	Valeur du logement	0,90	0,67	0,91	1,00	1,05	1,12	1,18
	Condominium enregistré	0,63	0,18	0,59	0,84	0,93	1,11	1,30
	Partie							
	Ne fait pas partie	0,15	0,07	0,14	0,19	0,28	0,39	0,47
	unifamiliaux	0,22	0,05	0,20	0,27	0,33	0,36	0,40
	Sans personnes additionnelles							
	Avec personnes additionnelles	0,50	0,20	0,48	0,61	0,72	0,74	0,79

Caractéristiques	Facteur	Centiles des facteurs au niveau de la RP					
		national ou provincial	1	50	75	90	95
Caractéristiques de la population							
Nombre de personnes dans la famille	0,18	0,08	0,19	0,24	0,33	0,41	0,45
économique							
Âge du conjoint, de la conjointe ou de la personne de référence dans la famille	1,42	0,80	1,37	1,53	1,60	1,78	1,91
économique							
Toutes les autres caractéristiques de la population	1,00	-	-	-	-	-	-
Caractéristiques des ménages et des logements							
Type de construction	0,33	0,05	0,35	0,55	0,67	0,75	0,89
Logement individuel non attenant							
Appartement de moins de 5 étages	0,57	0,12	0,56	0,70	0,83	0,99	1,26
Autres	0,91	0,18	0,88	0,99	1,18	1,23	1,32
Mode d'occupation	0,00	-	-	-	-	-	-
Période de construction	0,78	0,61	0,75	0,82	0,89	0,99	1,24
Principal type de chauffage/Principal combustible utilisé pour le chauffage	0,87	0,18	0,86	1,04	1,12	1,25	1,32
Chauffage central	0,42	0,09	0,38	0,54	0,60	0,70	0,89
Avec							
Sans	0,78	0,23	0,79	0,91	1,03	1,12	1,20
Taille du ménage							
Ménage d'une personne	0,00	-	-	-	-	-	-
Autres	0,76	0,19	0,72	1,09	1,17	1,21	1,30

Caractéristiques	Facteur	Centiles des facteurs au niveau de la RP						
		national ou provincial	1	50	75	90	95	99
Participation à la population active	Personnes occupées	0,75	0,59	0,76	0,83	0,86	0,91	0,93
	Personnes en chômage	1,06	0,76	1,04	1,14	1,20	1,27	1,38
	Personnes inactives	1,25	0,91	1,30	1,43	1,50	1,58	1,63
Principale source de revenu	Salaires et traitements	0,65	0,42	0,67	0,80	0,85	0,87	0,92
	Autres	1,05	0,71	1,00	1,12	1,17	1,20	1,24
	Incapacité	0,94	0,69	0,96	1,11	1,29	1,34	1,42
Limite(e) à la maison, à l'école et au travail	N'est pas limité(e)	0,61	0,41	0,58	0,69	0,74	0,78	0,81
	Situation des particuliers dans la famille de recensement	0,20	0,05	0,20	0,24	0,26	0,28	0,31
	Conjoint, conjointe, enfant	0,45	0,14	0,43	0,51	0,55	0,61	0,68
famille de recensement	Parent seul de sexe féminin	0,68	0,35	0,65	0,79	0,89	0,99	1,14
	Parent seul de sexe masculin, personne ne faisant pas partie d'une famille de recensement	0,14	0,06	0,16	0,21	0,28	0,34	0,36
	Parent seul, enfant	0,32	0,16	0,34	0,39	0,44	0,47	0,53
Autres membres de la famille	Nombre de personnes dans la famille de recensement	0,74	0,24	0,70	0,84	1,03	1,09	1,18
	Situation des particuliers dans la famille économique	0,14	0,06	0,16	0,21	0,28	0,34	0,36
	Conjoint, conjointe	0,32	0,16	0,34	0,39	0,44	0,47	0,53

Caractéristiques	Facteur	Centiles des facteurs au niveau de la RP					
		national ou provincial	1	50	75	90	95
Caractéristiques de la population							
Québec	1,15	0,18	1,10	1,51	1,76	1,81	1,99
Autres provinces	0,45	0,12	0,48	0,71	0,96	1,12	1,38
Canada	0,53	-	-	-	-	-	-
Langue maternelle – Français							
Québec	0,42	0,14	0,45	0,52	0,61	0,76	0,91
Nouveau-Brunswick	0,75	0,19	0,79	0,98	1,24	1,60	1,84
Autres provinces	1,04	0,09	1,12	1,49	1,71	1,89	2,06
Canada	0,77	-	-	-	-	-	-
Langue maternelle – Autres groupes linguistiques							
Langue maternelle – Autres groupes	1,70	0,73	1,63	2,11	2,44	2,51	2,60
Industrie/profession	0,92	0,25	0,80	1,13	1,25	1,31	1,38
Travail en 1985	0,89	0,62	0,92	1,14	1,22	1,29	1,31
Semaines travaillées en 1985	0,94	0,68	0,99	1,18	1,29	1,33	1,39
Heures travaillées pendant la semaine de référence	0,83	0,63	0,85	1,01	1,14	1,19	1,24
Année du dernier travail	0,89	0,60	0,94	0,99	1,05	1,11	1,20
En 1986, en 1985, avant 1985	1,18	0,80	1,15	1,34	1,43	1,50	1,67
Jamais travaillé	0,72	0,56	0,75	0,86	0,93	0,95	0,98
Travailleurs rémunérés	0,93	0,68	0,96	1,08	1,13	1,15	1,18
Travailleurs autonomes, entreprises non constituées en société, travailleurs familiaux non rémunérés							

Caractéristiques	Facteur	Centiles des facteurs au niveau de la RP				
		national ou provincial	1	50	75	90 95 99

Caractéristiques de la population						
Période d'immigration	Avant 1946, 1946-1966	0,98	1,02	1,10	1,22	1,37 1,45
	1967-1977, 1978-1982, 1983-1986	1,51	1,45	1,55	1,78	1,90 2,11
Âge à l'immigration	1,10	0,71	1,15	1,29	1,38	1,44 1,54
	Né(e) au Canada	1,09	0,82	1,08	1,16	1,18 1,20 1,21
	Né(e) à l'extérieur du Canada	1,35	1,11	1,34	1,43	1,60 1,67 1,75
Population immigrante/non immigrante						
Citoyenneté	Canada, par naissance	1,13	0,88	1,14	1,17	1,20 1,27 1,32
	Autres	1,59	1,04	1,40	1,65	1,88 1,95 2,12
Origine ethnique						
	Anglais, français	1,20	0,73	1,16	1,25	1,31 1,40 1,46
	Autres	1,65	1,07	1,57	1,70	1,89 1,99 2,11
Langue parlée à la maison						
	Anglais, français, anglais et français, anglais et langues non officielles	1,12	0,50	1,09	1,35	1,75 1,89 2,09
	Autres groupes linguistiques	1,76	0,99	1,68	1,89	2,01 2,20 2,41
Langues officielles						
	Anglais, français, anglais et français	1,05	0,69	1,01	1,18	1,31 1,42 1,58
	Autres groupes linguistiques	1,49	0,90	1,50	1,68	1,76 1,79 1,91
Langue maternelle – Anglais						
	Terre-Neuve, Île-du-Prince-Édouard, Nouvelle-Écosse, Colombie-Britannique	1,21	0,83	1,23	1,27	1,32 1,36 1,41
		0,92	0,24	0,96	1,45	1,62 1,90 2,23

Annexe D – Facteurs d'ajustement de l'erreur-type de 1986 au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération

Caractéristiques	Facteur	Centiles des facteurs au niveau de la RP				
		national ou provincial	1	50	75	90
Caractéristiques de la population	Groupes d'âge 0 à 4, 5 à 9, 10 à 14, 15 à 19, 20 à 24, 25 à 29	0,18	0,05	0,19	0,29	0,35
	Groupes d'âge 30 à 34, 35 à 44, 45 à 54, 55 à 59, 60 à 64, 65+, 15+	0,36	0,13	0,33	0,46	0,51
	Groupe d'âge 65+	0,00	-	-	-	-
	Sexe	0,00	-	-	-	-
	État matrimonial	0,25	0,04	0,23	0,31	0,42
	Célibataires, personnes mariées (sauf les personnes séparées)	0,88	0,55	0,84	0,98	1,06
Plus haut niveau de scolarité	Personnes séparées, divorcées, veufs et veuves	0,90	0,75	0,95	1,06	1,14
	Plus haut grade, certificat ou diplôme/Total des années de scolarité	1,20	0,84	1,16	1,22	1,28
	Principal domaine d'études	1,21	0,83	1,23	1,27	1,32
	Mobilité	1,61	0,90	1,60	1,75	1,85
Personnes n'ayant pas déménagé	Personnes n'ayant pas déménagé	1,41	0,97	1,36	1,41	1,41
	Personnes ayant déménagé (migrants, non-migrants)	2,09	1,41	1,41	1,41	1,41

Annexe C – Renseignements additionnels sur les statistiques utilisées dans l'Examen du biais d'échantillonnage

Que X représente la valeur connue d'une caractéristique 2A au niveau de la division de recensement (DR) et que $X^{(0)}$ représente l'estimateur de Horvitz-Thompson de X . $X^{(0)}$ a été calculé en multipliant le total-échantillon non pondéré pour la caractéristique de chaque SD échantillonné par l'inverse de la fraction de sondage des ménages réalisée pour le SD, puis en faisant la somme des résultats au niveau de la DR. Les SD non échantillonnés ont été exclus de l'analyse. L'écart-type de $X^{(0)}$, $\text{std}(X^{(0)}) = \sqrt{V(X^{(0)})}$ a été calculé selon l'hypothèse que des échantillons aléatoires simples de ménages ont été tirés indépendamment dans chaque SD (en fait, des échantillons aléatoires systématiques indépendants ont été tirés). Les variances ont donc été calculées au niveau des SD, puis on en a fait la somme au niveau de la DR. Les valeurs S^2 de la population ont été utilisées dans les calculs de la variance. Voir Cochran, 1977, pages 23-24, pour les formules de variance relatives aux caractéristiques des personnes et des familles, et pages 50-52 pour les formules de variance relatives aux caractéristiques des ménages et des logements.

Puisque les valeurs $X^{(0)}$ sont des estimateurs de Horvitz-Thompson, ils sont sans biais pour X . L'échantillonnage a été fait indépendamment dans différents SD. Par conséquent, les valeurs $X^{(0)}$ sont la somme de n variables aléatoires indépendantes, où n est le nombre de SD dans la DR. Puisque 90 % des DR comportent plus de 25 SD (la moyenne était de 140), n est très élevé dans la plupart des DR. Par conséquent, selon le théorème central limite, $Z^{(0)} = (X^{(0)} - X) / \text{std}(X^{(0)})$ devrait avoir une distribution approximativement normale $(0, 1)$ (voir Kendall et Stuart, 1963, p. 193). Ce ne serait pas le cas, toutefois, si les réponses 2B présentaient un biais important, pour quelque raison que ce soit.

Les valeurs $Z^{(0)}$ ont été produites pour l'ensemble des 281 DR échantillonnées au Canada, à l'égard des caractéristiques 2A données au chapitre V. Afin d'évaluer la normalité des valeurs $Z^{(0)}$ au niveau des DR, des histogrammes des valeurs $Z^{(0)}$ superposés à une FDP (fonction de densité de probabilité) normale ont été produits.

De plus, pour vérifier si $Z^{(0)}$ était tiré d'une distribution normale dont la moyenne est zéro (c.-à-d. si la méthode de sélection de l'échantillon était sans biais), on a calculé la moyenne $\bar{Z}^{(0)} = \sum_{i=1}^m Z_i^{(0)} / m$ où $m = 281$ (le nombre de DR) et

$Z_i^{(0)}$ est la valeur de $Z^{(0)}$ pour la i^{e} DR. En outre, l'écart-type des $Z_i^{(0)}$ a été déterminé, compte tenu que

$$\text{std}^2(Z^{(0)}) = \sum_{i=1}^m (Z_i^{(0)} - \bar{Z}^{(0)})^2 / (m-1). \text{ Puis la statistique } T \text{ donnée par } T_z = \sqrt{m} \bar{Z}^{(0)} / \text{std}(Z^{(0)}) \text{ a été calculée. Si la}$$

méthode de sélection de l'échantillon était non biaisée, T devrait suivre la distribution t de Student avec $m-1$ degrés de liberté. La probabilité que $|T| > 1.960$ si la méthode de sélection de l'échantillon était non biaisée est inférieure à 0,05. Donc, si $|T| > 1.960$, l'hypothèse que la méthode de sélection de l'échantillon était non biaisée sera rejetée et la différence entre l'estimation-échantillon et le chiffre de population sera dite statistiquement significative au niveau de 5 %.

Annexe B – Contraintes au niveau des RP et des SD appliquées aux coefficients de pondération du recensement de 1996

(Nota : Les contraintes ajoutées lors du deuxième exercice effectué en 1996 sont indiquées par le symbole « # ».)

Contraintes au niveau des SD

- Total des ménages dans le SD
- Total des personnes dans le SD

Contraintes au niveau des RP (personnes)

- Total des personnes
- Total des personnes âgées de ≥ 15 ans

Contraintes au niveau des RP (ménages)

- Ménages d'une personne #
- Ménages de deux personnes
- Ménages de trois personnes
- Ménages de quatre personnes
- Ménages de cinq personnes

- Hommes
- Hommes âgés de ≥ 15 ans

- Personnes âgées de 0 à 4 ans
- Personnes âgées de 5 à 9 ans
- Personnes âgées de 10 à 14 ans
- Personnes âgées de 15 à 19 ans
- Personnes âgées de 20 à 24 ans
- Personnes âgées de 25 à 29 ans
- Personnes âgées de 30 à 34 ans
- Personnes âgées de 35 à 39 ans
- Personnes âgées de 40 à 44 ans
- Personnes âgées de 45 à 49 ans
- Personnes âgées de 50 à 54 ans
- Personnes âgées de 55 à 59 ans
- Personnes âgées de 60 à 64 ans #
- Personnes âgées de 65 à 74 ans
- Personnes âgées de ≥ 75 ans

- Personnes mariées
- Personnes célibataires
- Personnes divorcées
- Veufs et veuves
- Union libre #

Annexe A – Glossaire

Les définitions des termes, variables et concepts du recensement sont présentées ici telles qu'elles figurent dans le *Dictionnaire du recensement de 1996* (n° 92-351-XPF au catalogue). L'utilisateur doit se reporter au *Dictionnaire du recensement de 1996* pour obtenir les définitions complètes et des observations additionnelles se rapportant aux concepts, comme des renseignements sur les variables directes et dérivées, ainsi que leur univers respectif.

Division de recensement (DR) : Terme générique qui désigne les régions géographiques établies en vertu de lois provinciales et qui sont des régions géographiques intermédiaires entre la municipalité (subdivision de recensement) et la province. Il peut s'agir de comtés, de districts régionaux, de municipalités régionales et d'autres genres de régions créées en vertu des lois provinciales.

État matrimonial : Situation conjugale d'une personne : marié(e) (y compris l'union libre); séparé(e); divorcé(e); veuf ou veuve; célibataire (jamais marié[e]).

Logement privé : Ensemble distinct de pièces d'habitation ayant une entrée privée donnant sur l'extérieur ou sur un corridor, un hall, un vestibule ou un escalier commun à l'intérieur. L'entrée doit donner accès au logement sans que l'on ait à passer par les pièces d'habitation de quelqu'un d'autre.

Logement privé occupé : Logement privé occupé de façon permanente par une personne ou un groupe de personnes. Sont également inclus dans cette catégorie les logements privés dont les résidents habituels sont temporairement absents le jour du recensement.

Ménage : Personne ou groupe de personnes (autres que des résidents étrangers) occupant un même logement et n'ayant pas de domicile habituel ailleurs au Canada. Il peut se composer d'un groupe familial (famille de recensement), avec ou sans autres personnes hors famille de recensement, de deux familles ou plus partageant le même logement, d'un groupe de personnes non apparentées ou d'une personne seule.

Ménage privé : Personne ou groupe de personnes (autres que des résidents étrangers) occupant un logement privé et n'ayant pas de domicile habituel ailleurs au Canada.

Secteur de dénombrement (SD) : Correspond à la région géographique dénombrée par un recenseur. Il s'agit de la plus petite région géographique normalisée pour laquelle des données du recensement sont établies. Tout le territoire du Canada est divisé en SD.

Secteur de recensement (SR) : Petites unités géographiques représentant des collectivités urbaines ou rurales créées au sein de régions métropolitaines de recensement et d'agglomérations de recensement dont le noyau urbanisé compte 50 000 habitants ou plus d'après le recensement précédent.

Subdivision de recensement (SDR) : Terme générique qui désigne les municipalités (telles que définies par les lois provinciales) ou leurs équivalents (par exemple, les réserves indiennes, les établissements indiens et les territoires non organisés).

ANNEXES

IX. Conclusion

L'échantillonnage fait maintenant partie intégrante du processus de recensement. Son utilisation permet de faire des économies substantielles et de réduire de façon appréciable le fardeau du répondant, ou encore, d'élargir la portée d'un recensement sans frais supplémentaires. Le prix de ces avantages est l'introduction d'une erreur d'échantillonnage dans les estimations obtenues à partir de l'échantillon. Cet effet se fait particulièrement sentir sur les estimations de faible valeur du recensement, qu'il s'agisse des chiffres se rapportant à des catégories peu usuelles au niveau provincial ou provincial, ou de ceux qui correspondent à des régions géographiques de petite taille. Il convient de souligner que les erreurs de réponse et les erreurs de dépouillement contribuent également à l'erreur totale des estimations du recensement et, là encore, ce sont les estimations de faible valeur qui sont particulièrement sensibles aux effets de ces erreurs non dues à l'échantillonnage. Par conséquent, bon nombre de ces faibles valeurs n'auraient qu'une fiabilité relative même si elles étaient fondées sur un recensement intégral. Pour ce qui est du recensement de 1996, les estimations inférieures ou égales à 50 établies à partir des données-échantillon sont généralement très peu fiables, tandis que les estimations allant jusqu'à 500 tendent à avoir des erreurs-types de plus de 10 % de leur valeur.

Pour de nombreuses caractéristiques, un certain biais a été détecté dans l'échantillon. Il a été établi que ce biais avait été causé partiellement par le traitement des données et à l'étape du contrôle et de l'imputation. Le reste du biais provient sans doute d'un ou plusieurs facteurs tels que le biais de non-réponse, le biais de réponse ou la sélection par les recenseurs d'un échantillon biaisé. Les méthodes utilisées pour pondérer les données-échantillon en vue d'obtenir des estimations à l'échelle de la population totale ont été appliquées avec succès et, de façon générale, ont permis d'atteindre les niveaux prévus de cohérences des estimations-échantillon et des chiffres de population. Bien que la cohérence qui a été obtenue aux niveaux provincial et national soit meilleure qu'en 1991, elle est quand même plus faible que ce à quoi on aurait pu s'attendre, compte tenu de la cohérence améliorée atteinte lors des tests pour les plus petits niveaux géographiques. Cela s'explique probablement par les petits biais accumulés dans l'échantillon sur de nombreuses années.

La méthode d'estimation du recensement sera réévaluée en vue du recensement de 2001, afin de voir s'il est possible d'améliorer la cohérence entre les estimations-échantillon et les chiffres de population au niveau des provinces et du Canada, tout en maintenant une bonne cohérence au niveau des SD. On devrait pouvoir obtenir en même temps des estimations plus fiables de l'erreur quadratique moyenne des estimations du recensement.

souffrent peut-être de biais semblables, mais aucune étude n'a été faite à ce sujet.

que dans celui de la MIQ. On peut le voir au tableau 6.5 du chapitre VI, où la PEGMC présente des différences population/estimation plus faibles que la MIQ pour les plus petites régions géographiques. Cependant, à mesure que la région géographique grossit, l'amélioration observée n'est plus aussi importante, et en 1991 la méthode itérative du quotient donnait des résultats supérieurs à ceux de la PEGMC au niveau provincial. Outre le biais introduit par l'échantillonnage et le traitement, Bankier, Rathwell et Majkowski, 1992, montrent, par une étude de Monte Carlo, que l'estimateur PEGMC est lui-même biaisé, bien que le biais relatif soit inférieur à 1 % pour 50 % des caractéristiques étudiées. Fait plus important, toutefois, la variance estimée des estimateurs PEGMC affiche un biais relatif médian de -25 % au niveau des RP. Ces estimateurs ont donc tendance à sous-estimer la variance vraie. Les estimateurs MIQ

Nombre total estimatif de personnes, de ménages et de logements					
Total estimé dans la région	500 000	1 000 000	2 500 000	5 000 000	10 000 000
50	15	15	15	15	15
100	20	20	20	20	20
250	30	30	30	30	30
500	45	45	45	45	45
1 000	65	65	65	65	65
2 500	100	100	100	100	100
5 000	140	140	140	140	140
10 000	200	200	200	200	200
25 000	310	310	310	320	320
50 000	420	440	440	450	450
100 000	570	600	620	630	630
250,000	710	870	950	970	990
500 000	0	1 000	1 260	1 340	1 380
1 000,000	0	0	1 550	1 790	1 900
2 500 000			0	2 240	2 740
5 000 000				0	3 160
10 000 000					4 900

Pour estimer l'erreur-type d'une estimation-échantillon donnée, le facteur d'ajustement s'appliquant à la caractéristique en question a été déterminé d'après l'annexe D. Le facteur d'ajustement au niveau national ou provincial pour les caractéristiques-échantillon se situait généralement entre 0,40 et 1,60. Ce facteur a ensuite été multiplié par l'erreur-type non ajustée choisie à partir du tableau 8.1.

L'exemple qui suit illustre la façon de calculer l'erreur-type ajustée. Supposons que l'estimation étudiée soit la population immigrante en Ontario. L'estimation du recensement de 1986 pour cette caractéristique était 2 081 200. Le chiffre de population obtenu pour l'Ontario au recensement de 1986 était 9 001 170. Si l'on utilise l'équation (4), qui permet de calculer l'erreur-type non ajustée, on obtient une estimation de 2 530. D'après l'annexe D, le facteur d'ajustement au niveau provincial pour la caractéristique « immigrante » est 1,12. Il s'ensuit que l'erreur-type ajustée pour cette estimation est $2\,530 \times 1,12 = 2\,834$.

Un deuxième exemple, toutefois, soulève un doute quant à l'exactitude de ces erreurs-types ajustées à titre d'estimations de la racine carrée de l'EQM. L'estimation, dans le recensement de 1986, du nombre de personnes avec « État matrimonial – Mariée(e) » vivant dans des logements privés dans les SD échantillonnées était 11 771 126. Le nombre de personnes, lors du recensement de 1986, qui vivaient dans des logements privés dans des SD échantillonnées était 11 778 842. La différence entre l'estimation et le chiffre de population est -7 716. Le quotient entre cette différence et l'erreur-type ajustée est $-7\,716 / 1\,233 = -6,25$. Un intervalle de confiance à 95 % pour une estimation serait normalement défini comme étant plus ou moins deux fois l'erreur-type ajustée. Le fait que le quotient entre la différence et l'erreur-type soit de -6,25 laisse entendre que l'erreur-type ajustée de 1 233 est une sous-estimation de la racine carrée de l'EQM.

B. Variance d'échantillonnage et biais dans l'estimation généralisée des moindres carrés

Dans Bankier, Rathwell et Majkowski, 1992, les coefficients de variation (CV) de la PEGMC pour certaines caractéristiques-échantillon ont été comparés aux CV correspondants de la MIQ. Dans les deux cas, les données du recensement de 1986 ont été utilisées. Le CV d'une estimation est la racine carrée de la variance estimée exprimée en pourcentage de l'estimation. Pour 79 RP, on a estimé les CV pour les estimations de 507 caractéristiques-échantillon au niveau des SD et 642 caractéristiques-échantillon au niveau des RP (chacune d'entre elles s'appliquant à un total minimum estimé de 60 ménages dans la population). Chacune des estimations au niveau des RP et au niveau des SD a été classée comme estimation faible (inférieure ou égale à la médiane des estimations) ou comme estimation élevée (supérieure à la médiane des estimations). On a constaté que la valeur médiane des CV pour les estimations élevées au niveau des RP était de 5 % pour la PEGMC, tandis qu'elle était de 6 % pour la MIQ. La valeur médiane des CV pour les estimations faibles au niveau des RP était de 13 % pour la PEGMC et de 15 % pour la MIQ. La valeur médiane des CV pour les estimations élevées au niveau des SD était de 10 % pour la PEGMC et de 12,5 % pour la MIQ. La PEGMC a donc permis une réduction des CV par rapport à la MIQ tant au niveau des SD qu'à celui des RP. Puisque les variations à des niveaux géographiques supérieurs sont simplement la somme des variances au niveau des RP, ces réductions des CV devraient se répercuter sur les niveaux géographiques plus élevés.

Il a été indiqué au chapitre V que l'échantillon du recensement affichait des biais faibles, mais significatifs. Ces biais ne sont pas significatifs par rapport à la variance d'échantillonnage au niveau des RP. À des niveaux géographiques supérieurs, toutefois, le biais touchant une caractéristique peut s'accumuler s'il produit systématiquement des

d'évaluer les effets combinés du plan de sondage et de la méthode d'estimation sur la variance d'échantillonnage. On peut ensuite ajuster les estimations simples de la variance d'échantillonnage, qui sont peu coûteuses à produire, pour tenir compte de cet effet et produire une estimation de la variance d'échantillonnage pour n'importe quelle estimation du recensement.

Le tableau 8.1 donne des valeurs non ajustées (simples) de l'erreur-type des estimations-échantillon du recensement. Les valeurs présentées dans ce tableau ont été établies en supposant qu'on a utilisé un échantillon aléatoire simple avec fraction de sondage de 1 sur 5, et une pondération simple de 5. Les erreurs-types sont présentées au tableau 8.1 en fonction à la fois de la taille de l'estimation du recensement et de la taille de la région géographique. Par exemple, pour une estimation de 50 personnes dans une région géographique comptant 500 personnes, l'erreur-type non ajustée est de 15.

Le tableau 8.1 contient les erreurs-types associées à un nombre restreint de valeurs du « total estimé » et du nombre total de personnes, de ménages, de logements ou de familles dans la région. La formule qui suit peut servir à calculer l'erreur-type non ajustée de n'importe quel total estimé, pour une région de n'importe quelle taille :

$$ETNA = \sqrt{\frac{4E(N-E)}{N}} \quad (4)$$

où ETNA est l'erreur-type non ajustée, E est le total estimé et N est le nombre total de personnes, de ménages, de logements ou de familles dans la région. Par exemple, pour un total estimé de 750 personnes dans une région comptant 9 000 personnes, l'erreur-type non ajustée serait de :

$$\sqrt{\frac{4(750)(9,000-750)}{9,000}} = 52$$

L'Étude sur la variance d'échantillonnage de 1986 fournit des facteurs d'ajustement⁹ par lesquels on doit multiplier les erreurs-types non ajustées pour tenir compte des effets combinés du plan de sondage et de la méthode d'estimation. Pour calculer ces facteurs d'ajustement, un échantillon de 401 RP (sur un total de 5 941 RP) a été tiré. On a réparti l'échantillon parmi les 10 provinces¹⁰ de manière à obtenir de bonnes estimations de la variance d'échantillonnage au niveau provincial sans pour autant compromettre indûment la qualité des estimations au niveau national. Pour chaque RP dans l'échantillon, les estimations de la variance d'échantillonnage des **estimations produites par la MIQ** ont été calculées pour différentes catégories de toutes les caractéristiques énumérées au **tableau 9 du Guide à l'intention des utilisateurs du recensement de 1986** : ce tableau est inclus dans l'annexe D. (NOTA IMPORTANTE : Ces facteurs ont été calculés à partir des données du recensement de 1996, alors que l'on utilisait la méthode itérative du quotient; si l'on utilise ce tableau, on suppose que l'incidence des ajustements est la même pour le recensement de 1996, même si une méthode d'estimation différente a été utilisée.) Les estimations de la variance d'échantillonnage aux niveaux provincial et national ont été obtenues en pondérant les estimations au niveau de la RP. Les facteurs d'ajustement pour chaque sous-catégorie de chacune des caractéristiques ont été calculés en divisant la racine carrée de ces estimations par l'erreur-type non ajustée correspondante. Des facteurs d'ajustement ont été calculés aux niveaux provincial et national pour chaque caractéristique en faisant la moyenne des facteurs d'ajustement de toutes les catégories. Pour de plus amples renseignements sur la façon dont ces facteurs d'ajustement ont été calculés, voir Bédard, 1990.

⁹ On appelle communément « effet du plan » le carré des facteurs d'ajustement.

¹⁰ Le territoire du Yukon et les Territoires du Nord-Ouest ont été groupés avec la Colombie-Britannique.

VIII. Variance d'échantillonnage

L'erreur d'échantillonnage a deux composantes : la variance et le biais. La variance mesure la variabilité de l'estimation par rapport à l'estimation moyenne qui résulterait de répétitions hypothétiques du processus d'enquête. Le biais est la différence entre la valeur moyenne d'une estimation qui résulterait de répétitions hypothétiques du processus d'enquête, et la valeur vraie de la variable estimée. L'erreur quadratique moyenne (EQM) mesure la variabilité de l'estimation par rapport à la valeur vraie, pour des répétitions hypothétiques du processus d'enquête. On peut montrer que l'EQM est égale à la somme de la variance et du carré du biais. L'EQM est la mesure la plus exacte de l'ampleur moyenne de l'écart entre l'estimation et la valeur vraie de la population. Si le biais est faible par rapport à la variance, la variance est une bonne approximation de l'EQM. On constate toutefois qu'en réalité, le biais s'accumule à mesure qu'on produit des estimations du recensement visant des régions géographiques plus grandes. Le biais peut donc être non significatif pour de petites régions géographiques, mais devenir élevé par rapport à la variance pour de grandes régions géographiques. Il est possible, par conséquent, que la variance soit beaucoup plus faible que l'EQM pour de grandes régions géographiques. On peut estimer la variance d'une estimation à partir de l'échantillon, mais on ne peut pas faire de même dans le cas du biais d'une estimation. Autrement dit, il n'est pas possible d'estimer avec exactitude l'EQM à partir de l'échantillon, à moins que le biais soit faible par rapport à la variance.

Dans les recensements précédents, une étude visant à produire des estimations de la variance d'échantillonnage a été réalisée. La section A donne quelques résultats de l'étude de 1986 (pour de plus amples renseignements, voir le *Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération*). Comme au niveau des grandes régions géographiques, le biais semble être l'élément dominant de l'EQM (voir le chapitre V), le calcul de la variance d'échantillonnage ne donne pas une estimation exacte de l'EQM pour les grandes régions géographiques, c'est pourquoi il a été décidé de ne pas répéter cette étude pour le recensement de 1996. On trouvera à la section B, cependant, quelques observations relatives à l'impact qu'a eu la méthode d'estimation utilisée dans le recensement de 1996 sur la variance d'échantillonnage, par rapport au recensement de 1986.

A. Étude sur la variance d'échantillonnage du recensement de 1986

Le chapitre V présente les résultats de l'Examen du biais d'échantillonnage et décrit la nature et l'importance du biais dans l'échantillon avant la pondération. Les chapitres VI et VII présentent les résultats ayant trait au biais d'échantillonnage après la pondération. Même avec une méthode d'échantillonnage totalement dépourvue de biais, les résultats seraient sujets à la variance du simple fait que les estimations ont été produites à partir d'un échantillon. La variance peut être estimée à l'aide des données recueillies dans l'enquête-échantillon⁸. L'Étude de la variance d'échantillonnage de 1986 a été effectuée afin d'estimer l'effet des méthodes d'échantillonnage et d'estimation sur les chiffres du recensement établis à partir de données-échantillon.

Statistique Canada produit des milliers de tableaux à partir des données-échantillon (c'est-à-dire celles recueillies sur la formule 2B). Théoriquement, on pourrait fournir une mesure de précision (à savoir l'estimation de la variance d'échantillonnage) pour chaque estimation présentée dans ces tableaux. Cette mesure tiendrait compte à la fois du plan de sondage et de la méthode d'estimation. En pratique cependant, il est impossible de calculer une telle mesure pour toutes les estimations du recensement en raison du coût élevé du traitement des données. On calcule donc la variance d'échantillonnage pour un sous-ensemble seulement des estimations du recensement. Cela permet ensuite

⁸ Malheureusement, la variance d'échantillonnage ne nous fournit aucune indication de l'importance de l'erreur non due à l'échantillonnage.

Tableau 7.4 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SD - recensements de 1996 et de 1991

Caractéristiques étudiées	Centiles de 1996					Centiles de 1991				
	10e	25e	50e	75e	90e	10e	25e	50e	75e	90e
Personnes										
Hommes	-4,97	-2,05	0,02	2,14	4,80	-6,25	-2,59	0,07	2,64	6,06
Femmes	-4,91	-2,08	0	1,96	4,77	-6,29	-2,66	-0,10	2,55	6,24
Population totale	-1,94	0	0	0	1,81	-3,56	0	0	0	3,35
Âge 0 à 4	-21,40	-10,30	0	10,60	21,78	-26,40	-12,70	-0,22	12,50	26,49
Âge 5 à 9	-17,30	-8,68	-0,10	8,63	17,87	-24,20	-11,50	0,03	11,23	24,00
Âge 10 à 14	-19,20	-9,33	-0,15	9,63	20,01	-25,40	-12,20	0,07	12,23	25,28
Âge 15 à 19	-16,70	-8,26	0,11	8,94	17,86	-25,40	-12,10	0,37	13,15	27,12
Âge 20 à 24	-18,30	-9,30	-0,12	9,39	19,20	-27,90	-13,60	-0,23	13,51	28,48
Âge 25 à 29	-18,40	-9,44	-0,32	9,42	19,36	-26,10	-12,60	-0,51	11,45	25,23
Âge 30 à 34	-17,10	-8,73	-0,14	8,56	17,72	-24,40	-11,60	-0,23	11,43	24,63
Âge 35 à 39	-16,90	-8,53	-0,30	8,22	17,14	-25,10	-12,40	-0,68	11,28	24,60
Âge 40 à 44	-17,20	-9,02	-0,34	8,52	17,92	-26,30	-12,80	-0,37	11,93	26,53
Âge 45 à 49	-17,30	-9,02	-0,40	8,60	18,39	-27,40	-13,10	0	13,65	29,16
Âge 50 à 54	-19,00	-9,59	-0,54	9,37	19,54	-30,00	-15,20	-0,10	14,85	31,77
Âge 55 à 59	-19,90	-10,60	-0,74	9,96	20,66	-29,60	-15,00	0,29	15,76	31,40
Âge 60 à 64	-20,50	-10,30	0,06	10,82	22,07	-30,30	-15,40	-0,50	16,01	31,98
Âge 65 à 74	-17,30	-9,12	-0,40	8,60	17,73	-26,60	-12,60	-0,63	11,72	27,41
Âge 75 et plus	-18,20	-8,72	-0,06	8,78	18,10	-27,50	-12,70	-0,32	12,50	27,83
Personnes célibataires	-6,74	-2,76	-0,39	2,84	6,22	-8,45	-3,43	0,04	3,42	8,03
Personnes mariées	-7,85	-3,39	0,06	3,41	8,08	-8,78	-3,55	0,07	3,76	8,92
Veufs et veuves	-18,00	-9,39	0,10	8,69	17,55	-23,50	-11,50	-0,60	10,68	22,99
Personnes divorcées	-18,90	-9,61	0	9,86	19,77	-29,70	-15,20	-0,93	14,50	29,84
Personnes séparées	-33,10	-20,80	-0,36	11,74	29,78	-28,90	-11,70	3,21	17,38	32,19
Union libre = oui	-6,00	0	0	0	6,66	-	-	-	-	-
Ménages										
Ménages d'une personne	-13,10	-6,01	0,01	6,00	12,62	-13,50	-6,67	-0,35	5,69	12,38
Ménages de deux personnes	-13,80	-6,96	-0,23	6,61	13,67	-14,40	-7,08	0	7,06	15,11
Ménages de trois personnes	-16,00	-8,40	-0,17	7,88	16,21	-22,30	-11,40	-0,12	11,25	23,48
Ménages de quatre personnes	-13,90	-7,04	-0,19	7,07	14,20	-17,60	-8,41	0	8,49	18,45
Ménages de cinq personnes	-16,40	-8,23	-0,10	8,46	16,68	-22,20	-11,00	1,09	13,92	26,33
Ménages de six personnes ou plus	-13,30	-5,98	2,01	10,03	18,95	-22,60	-9,30	2,48	14,81	26,63

Caractéristiques étudiées	Centiles de 1996					Centiles de 1991				
	10e	25e	50e	75e	90e	10e	25e	50e	75e	90e
Personnes										
Hommes	-0,20	0	0	0	0,26	-0,90	0	0	0	0,87
Femmes	-0,20	0	0	0	0,25	-0,90	0	0	0	0,79
Population totale	0	0	0	0	0	-0,20	0	0	0	0,24
Âge 0 à 4	-1,40	0	0	0	1,80	-11,00	-1,20	0	1,03	9,77
Âge 5 à 9	-1,90	0	0	0	1,72	-11,00	-1,60	0	0,94	10,20
Âge 10 à 14	-1,90	0	0	0	1,72	-11,00	-1,30	0	2,23	13,00
Âge 15 à 19	-3,10	0	0	0	3,62	-11,00	-2,70	0	3,93	11,60
Âge 20 à 24	-2,10	0	0	0	2,44	-10,00	-1,50	0	1,90	10,80
Âge 25 à 29	-1,90	0	0	0	1,89	-10,00	-1,60	0	0,28	7,31
Âge 30 à 34	-1,10	0	0	0	1,28	-7,80	-0,60	0	0,29	7,46
Âge 35 à 39	-1,00	0	0	0	1,16	-9,30	-1,30	0	0	7,88
Âge 40 à 44	-1,20	0	0	0	1,55	-10,00	-1,40	0	1,02	9,05
Âge 45 à 49	-1,10	0	0	0	1,64	-13,00	-3,10	0	3,06	12,00
Âge 50 à 54	-1,80	0	0	0	2,48	-15,00	-5,00	0	5,84	16,00
Âge 55 à 59	-3,60	0	0	0	3,62	-16,00	-5,00	0	5,84	17,30
Âge 60 à 64	-7,00	0	0	0	9,89	-17,00	-6,50	0	6,54	18,60
Âge 65 à 74	-2,70	0	0	0	2,34	-14,00	-3,00	0	1,61	12,40
Âge 75 et plus	-14,00	0	0	0	7,31	-22,00	-8,90	0	8,23	21,00
Personnes célibataires	-0,30	0	0	0	0,31	-1,20	0	0	0	1,27
Personnes mariées	-0,40	0	0	0	0,35	-1,40	0	0	0	1,59
Veufs et veuves	-4,40	0	0	0	3,36	-15,00	-5,30	0	3,87	14,90
Personnes divorcées	-2,20	0	0	0	2,97	-16,00	-5,20	0	4,08	14,20
Personnes séparées	-4,60	0	0	0	5,53	-24,00	-8,80	0	8,86	25,50
Union libre = oui	-2,50	0	0	0	3,39	-	-	-	-	-
Ménages										
Ménages d'une personne	-2,80	-0,80	0	0,29	2,06	-7,60	-3,20	-0,40	1,38	5,39
Ménages de deux personnes	-1,50	0	0	0	1,11	-3,90	-0,30	0	0,40	4,23
Ménages de trois personnes	-1,60	0	0	0	1,79	-8,90	-2,30	0	3,32	9,95
Ménages de quatre personnes	-0,90	0	0	0	2,01	-7,60	-0,90	0	2,19	9,90
Ménages de cinq personnes	-3,60	0	0	0,12	8,01	-16,00	-4,80	0	7,47	18,50
Ménages de six personnes ou plus	-6,00	-2,30	1,04	3,78	7,23	-24,00	-11,00	-0,50	7,11	16,30

recensements de 1996 et de 1991

Tableau 7.3 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SR –

Tableau 7.2 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SDR – recensements de 1996 et de 1991

Caractéristiques étudiées	Centiles de 1996					Centiles de 1991				
	10e	25e	50e	75e	90e	10e	25e	50e	75e	90e
Personnes										
Hommes	-4,27	-1,22	0	1,30	4,17	-5,47	-1,69	0	1,62	5,05
Femmes	-4,49	-1,33	0	1,22	4,27	-5,49	-1,83	0	1,83	5,56
Population totale	-2,11	0	0	0	1,94	-3,36	-0,37	0	0,26	3,28
Âge 0 à 4	-16,40	-4,20	0	3,53	15,47	-19,40	-7,14	0	6,23	18,50
Âge 5 à 9	-13,20	-3,45	0	3,66	13,98	-16,50	-6,11	0	5,73	17,01
Âge 10 à 14	-13,00	-4,19	0	3,91	13,02	-17,00	-6,15	0	7,06	18,01
Âge 15 à 19	-12,90	-3,20	0	3,87	13,64	-16,80	-6,03	0	6,92	19,04
Âge 20 à 24	-14,10	-3,97	0	3,47	14,06	-21,00	-7,71	0	7,94	21,04
Âge 25 à 29	-14,20	-3,27	0	4,03	15,15	-20,30	-7,54	0	5,96	19,29
Âge 30 à 34	-12,70	-3,40	0	3,99	12,94	-17,90	-6,21	0	6,02	17,39
Âge 35 à 39	-13,00	-3,43	0	3,70	13,23	-19,20	-6,66	0	5,86	18,71
Âge 40 à 44	-13,50	-3,70	0	3,49	12,81	-19,20	-6,43	0	7,42	19,92
Âge 45 à 49	-13,00	-3,27	0	3,90	13,81	-19,20	-7,37	0	8,45	22,15
Âge 50 à 54	-14,60	-3,46	0	3,27	13,84	-23,10	-9,06	0	8,62	21,29
Âge 55 à 59	-14,80	-2,95	0	4,00	15,20	-22,30	-8,28	0	8,34	22,23
Âge 60 à 64	-15,70	-3,73	0	4,79	15,54	-22,30	-9,43	0	9,14	22,91
Âge 65 à 74	-14,80	-4,25	0	3,91	13,33	-21,20	-8,49	0	6,72	19,36
Âge 75 et plus	-18,40	-5,91	0	4,14	15,11	-26,20	-12,00	-1,31	7,74	22,02
Personnes célibataires	-5,46	-1,61	0	1,83	5,46	-7,26	-2,34	0	2,33	7,14
Personnes mariées	-5,83	-1,83	0	1,73	5,64	-6,33	-1,99	0	2,31	6,98
Veufs et veuves	-14,40	-4,13	0	3,55	14,45	-18,90	-8,35	0	6,03	17,18
Personnes divorcées	-14,90	-2,73	0	3,20	15,10	-19,60	-7,28	0	7,72	19,43
Personnes séparées	-11,10	-0,82	0	0,88	10,51	-20,50	-7,85	0	8,55	22,51
Union libre = oui	-17,40	-4,84	0	3,96	16,12	-	-	-	-	-
Ménages										
Ménages d'une personne	-11,10	-3,16	-0,14	2,87	10,48	-11,90	-5,07	-0,54	3,51	10,46
Ménages de deux personnes	-10,80	-3,29	0	3,15	10,09	-11,10	-3,71	0	4,02	11,56
Ménages de trois personnes	-10,60	-2,56	0	2,82	11,37	-15,30	-5,54	0	6,18	17,21
Ménages de quatre personnes	-10,20	-2,93	0	2,60	10,11	-14,80	-5,03	0	4,44	13,50
Ménages de cinq personnes	-7,89	-0,91	0	3,69	11,42	-14,00	-4,94	0,40	8,11	19,45
Ménages de six personnes ou plus	-9,36	-3,29	-0,28	2,12	5,84	-20,80	-10,40	-3,38	4,13	12,07

Tableau 7.1 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les DR – recensements de 1996 et de 1991

[illegible]

¹ Comme la caractéristique « Union libre = oui » n'a pas été utilisée comme contrainte en 1991, il est impossible d'établir une comparaison entre 1996 et 1991.

et 75^e centiles correspondent à zéro pour presque toutes les caractéristiques, contrairement aux intervalles obtenus en 1991. De plus, les intervalles entre le 10^e et le 90^e centiles sont beaucoup plus faibles en 1996. De nouvelles spécifications ont été utilisées pour la formation des RP en 1996. Elles ont permis de réduire le nombre de cas de non-respect des limites de SR en regard de 1991. Ce changement pourrait contribuer à expliquer les améliorations notées en 1996.

D. Secteurs de dénombrement (SD)

Les SD sont les composantes des RP, et les RP sont le niveau le plus bas auquel les estimations-échantillon sont forcées à correspondre aux chiffres de population pour la plupart des caractéristiques. Les SD sont aussi les composantes de niveaux géographiques plus élevés (DR, SDR, SR, etc.) et plusieurs des RP, comme le montre le tableau 7.3, sont des composantes de ces niveaux supérieurs. Par conséquent, on ne peut pas s'attendre à ce que la cohérence au niveau des SD soit aussi bonne qu'aux niveaux géographiques plus élevés qui ont été étudiés. Le tableau 7.4 confirme que, pour la plupart des caractéristiques étudiées, dans les SD échantillonnées ayant un chiffre de population pour la caractéristique supérieure à 50, les écarts sont supérieurs à ceux observés aux niveaux géographiques examinés précédemment. C'est le cas aussi bien en 1996 qu'en 1991. Par rapport aux écarts de 1991 aux 10^e et 90^e centiles (et aux 25^e et 75^e centiles), les écarts de 1996 sont beaucoup plus faibles pour la plupart des caractéristiques étudiées. Les valeurs pour les ménages d'une personne sont cependant très semblables pour 1991 et 1996.

E. Incidence des changements apportés à la méthode de pondération en 1996

Il existe plusieurs explications possibles à l'augmentation du niveau de cohérence observé en 1996 par rapport à 1991 :

- On a utilisé un nombre beaucoup moins grand de contraintes en 1996, de sorte que les contraintes retranchées ont été moins nombreuses, ce qui a permis d'utiliser des ensembles de contraintes très similaires dans toutes les RP. Le degré de cohérence entre l'estimation et le chiffre de population pour une caractéristique donnée est par conséquent plus stable parmi les RP.
- En 1996, les coefficients de pondération calculés à l'aide de la méthode pouvaient être inférieurs à 1. Ce fait a également contribué à réduire le nombre de contraintes retranchées.

- Les modifications au chapitre de la formation des RP ont entraîné une réduction du nombre de cas de non-respect des limites des régions géographiques de niveau supérieur. Certaines grandes régions géographiques sont maintenant constituées de RP plus complètes, ce qui contribue à réduire les écarts dans ces régions.

En conclusion, il semble que les changements apportés à la méthode de pondération en 1996 ont permis d'obtenir de meilleures estimations qu'en 1991 pour ce qui est des petites régions géographiques car, dans les régions de même niveau géographique, l'écart associé à une caractéristique donnée est généralement plus près de 0 en 1996 qu'en 1991.

0 à 4 variait entre -0,44 % et 0,05 % pour 80 % des DR. Une répartition symétrique de part et d'autre de 0 signifie que les différences entre les estimations-échantillon et les chiffres de population sont réparties de façon égale. Cependant, une répartition non symétrique signifie que les chiffres de l'échantillon ne sont pas répartis de façon uniforme entre les différences positives et négatives.

Toutes les DR sont formées de RP complètes. Ainsi, parmi les caractéristiques qui se sont traduites en contraintes en 1996, celles qui étaient rarement ou jamais retranchées dans une RP affichaient une cohérence quasi parfaite au niveau des DR⁷. Ces caractéristiques étaient les suivantes : nombre total de personnes, nombre total de femmes, nombre total de célibataires, etc. Le niveau de cohérence observé pour les autres caractéristiques n'était pas parfait, mais demeurerait quand même très bon, sauf pour les caractéristiques représentant un faible pourcentage de la population dans la plupart des DR, par exemple, le nombre de personnes séparées ou le nombre de ménages de six personnes ou plus. Une relation générale existe entre les écarts et les chiffres de population pour toutes les caractéristiques, c'est-à-dire que la cohérence s'accroît à mesure qu'augmente le chiffre de population pour la DR.

Le tableau 7.1 montre également les centiles correspondant aux écarts de 1991 pour les DR. Les tableaux 7.2, 7.3 et 7.4 contiennent également les données de 1991 pour les autres niveaux géographiques. Si l'on compare les données de 1991 à celles de 1996, on constate que les écarts enregistrés en 1996 au niveau des DR sont identiques ou significativement inférieurs à ceux de 1991 pour toutes les caractéristiques figurant au tableau 7.1.

La taille des écarts au niveau des DR est très faible par comparaison avec les écarts observés aux autres niveaux géographiques plus petits, qui sont traités dans les sections qui suivent.

B. Subdivisions de recensement (SDR)

Le tableau 7.2 résume les résultats concernant la cohérence entre les estimations-échantillon et les chiffres de population pour toutes les SDR échantillonnées au Canada dont le chiffre de population pour la caractéristique était supérieure à 50. Il porte sur les mêmes caractéristiques que le tableau 7.1. Les SDR ne sont pas toujours uniquement formées de RP entières. En outre, elles sont beaucoup plus petites en moyenne que les DR. Par conséquent, le niveau de cohérence n'est pas aussi bon pour les SDR que pour les DR. De façon générale, comme dans le cas des DR, la cohérence devenait meilleure à mesure qu'augmentaient les chiffres de population dans les SDR, et ce, pour toutes les caractéristiques. Comparativement aux écarts de 1991 entre le 10^e et le 90^e centile, les écarts de 1996 sont beaucoup plus faibles pour presque toutes les caractéristiques. Ceci s'avère également vrai pour l'intervalle entre le 25^e et le 75^e centile.

C. Secteurs de recensement (SR)

Le tableau 7.3 présente un résumé du niveau de cohérence pour tous les SR échantillonnés au Canada. Comme dans le cas des tableaux 7.1 et 7.2, ce tableau comprend uniquement les SR où les chiffres de population associés à la caractéristique étudiée étaient supérieurs à 50. Les SR ont en moyenne des populations plus importantes que les SDR. L'augmentation du niveau de cohérence au niveau des SR entre 1991 et 1996 est très marquée. En 1996, les 25^e

⁷ Même lorsque les caractéristiques présentent une cohérence parfaite, les totalisations publiées des caractéristiques de base établies à partir des données-échantillon ne concordent pas parfaitement avec les totalisations des mêmes caractéristiques basées sur les données intégrales. Cette différence est attribuable au fait que les résidents des logements collectifs qui n'avaient pas à répondre aux questions-échantillon (voir le chapitre II, section B) sont inclus dans les totalisations faites à partir des données intégrales, mais exclus des totalisations fondées sur les données-échantillon.

VII. Examen de la cohérence entre les estimations-échantillon et les chiffres de population

Pour que la PEGMC soit efficace, il a fallu retrancher certaines contraintes dans chaque RP avant de calculer les coefficients de pondération. Plusieurs caractéristiques importantes ont donc été exclues dans un certain nombre de RP. Cela a eu pour conséquence de réduire le niveau de concordance (cohérence) entre les estimations-échantillon et les chiffres de population pour ces caractéristiques. De plus, de nombreuses régions géographiques d'intérêt ne se composent pas toujours de RP complètes. Dans ces régions, la cohérence pour l'ensemble des caractéristiques dépend donc de la mesure dans laquelle les régions se composent de RP complètes.

L'examen de la cohérence a porté sur les écarts entre les estimations-échantillon et les chiffres de population (exprimés en pourcentage des chiffres de population) pour le même ensemble de base de 31 caractéristiques que celles de l'Examen du biais d'échantillonnage (voir l'annexe B) pour les régions géographiques suivantes :

- a) divisions de recensement;
- b) subdivisions de recensement;
- c) secteurs de recensement;
- d) secteurs de dénombrement.

Comme il est indiqué au chapitre VI (sous-section B.2), les écarts entre les estimations-échantillon et les chiffres de population ont été calculés de la façon suivante :

$$\text{écart} = \frac{\text{Estimation-échantillon} - \text{chiffre de population}}{\text{chiffre de population}} \times 100$$

A. Divisions de recensement (DR)

Les centiles du tableau 7.1 résument les résultats relatifs au niveau de cohérence, pour l'ensemble des DR échantillonnées au Canada, d'une grande variété de caractéristiques de base dont le chiffre de population⁶ est supérieur à 50. En général, les écarts obtenus pour des caractéristiques dont le chiffre de population était de ≤ 50 dans la plupart des régions géographiques se sont révélés relativement élevés (soit positifs, soit négatifs). On a donc décidé de ne pas inclure les régions géographiques où le chiffre de population pour la caractéristique était inférieur ou égal à 50, car il suffisait de quelques-unes de ces régions pour perturber sensiblement les centiles des écarts présents dans les tableaux du présent chapitre. Cette perturbation se produirait si une grande partie de ces régions présentait soit des écarts positifs relativement élevés, soit des écarts négatifs relativement élevés.

Au tableau 7.1, pour chaque caractéristique, N % des DR présentaient des écarts inférieurs au N^e centile tandis que 100 – N % des DR avaient des écarts supérieurs au N^e centile. L'écart se situait donc entre le 10^e et le 90^e centile pour 80 % des DR, et entre le 25^e et le 75^e centile pour 50 % des DR, etc. Par exemple, l'écart pour le groupe d'âge

⁶ Il s'agit du chiffre de population pour la caractéristique. Par exemple, le niveau de cohérence pour le groupe d'âge 0 à 4 est résumé pour toutes les DR dans lesquelles plus de 50 personnes appartenaient au groupe d'âge 0 à 4. La même définition vaut pour les tableaux 7.1, 7.2, 7.3 et 7.4.

$$R = 100 * \frac{\sum_{N_{96}}^1 |X_{96} - \hat{X}_{96}| / \sum_{N_{86}}^1 |X_{96} - \hat{X}_{86}|}{\sum_{N_{96}}^1 |X_{96} - \hat{X}_{96}| / \sum_{N_{86}}^1 |X_{96} - \hat{X}_{86}|} \quad (3)$$

où X_{96} et X_{86} sont respectivement les chiffres de population de 1996 et 1986 pour une caractéristique. L'estimation-échantillon de 1996 d'après les coefficients de pondération PEGMC est \hat{X}_{96} , tandis que l'estimation-échantillon de 1986 fondée sur les coefficients de pondération MIQ est \hat{X}_{86} . Des valeurs R ont été calculées pour chacun des six niveaux géographiques (SD, RP, SDR, DR, province et Canada). La somme des valeurs absolues des différences population/estimation a été calculée, N_{96} étant égal au nombre d'unités du niveau géographique visé en 1996 et N_{86} étant égal au nombre d'unités du niveau géographique visé en 1991. Une valeur R se situant entre 95 et 105 indique que les systèmes d'estimation de 1996 et de 1991 ont donné à peu près le même résultat. Une valeur R inférieure à 95 indique que le système de 1996 a été plus efficace que celui de 1991 pour la caractéristique visée au niveau géographique en question, tandis qu'une valeur R supérieure à 105 indique qu'il a été moins efficace. Le tableau présente également les résultats de 1991, alors que la comparaison avait été établie pour 62 caractéristiques (dont 49 étaient des contraintes).

Tableau 6.6 Pourcentage des caractéristiques ayant des valeurs R dans certains intervalles

Valeur R	SD	RP	SDR	DR	Province	Canada
1996 vs 1986						
< 95	94	64	94	64	52	49
95 à 105	0	15	0	15	21	18
> 105	6	21	6	21	27	33
1991 vs 1986						
< 95	87	58	81	47	31	29
95 à 105	11	11	8	18	14	10
> 105	2	31	11	35	55	61

La méthode d'estimation utilisée en 1996 a permis de réduire les différences entre les chiffres de population et les estimations à tous les niveaux comparativement à la méthode de 1986. Il est à noter qu'en 1991 la méthode d'estimation a permis de réduire les différences au niveau des petites régions géographiques, du SD jusqu'à la DR, mais qu'aux niveaux provincial et national, la méthode de 1986 a donné de plus petites différences.

Tableau 6.5. Fréquence à laquelle les contraintes au niveau des RP ont été retranchées en 1996

Caractéristique	Faible	LD	QLD	Observation aberrante	Total
Hommes	0	0	0	1	1
Femmes*	-	-	-	-	-
Population totale	0	0	0	0	0
Âge 0 à 4	6	3 071	20	57	3 154
Âge 5 à 9	30	709	77	135	951
Âge 10 à 14	35	2 110	33	61	2 239
Âge 15 à 19	6	514	27	96	643
Âge 20 à 24	1	216	133	119	469
Âge 25 à 29	1	347	108	82	538
Âge 30 à 34	1	29	23	42	95
Âge 35 à 39	1	0	6	31	38
Âge 40 à 44	1	3	13	45	62
Âge 45 à 49	1	4	9	50	64
Âge 50 à 54	2	157	67	83	309
Âge 55 à 59	2	636	213	147	998
Âge 60 à 64	3	1 122	973	128	2 226
Âge 65 à 74	4	3	214	81	302
Âge 75 et plus	36	2 864	100	60	3 060
Personnes célibataires	0	0	1	3	4
Personnes mariées	0	0	0	4	4
Veufs et veuves	2	0	174	345	521
Personnes divorcées	1	1	213	252	467
Personnes séparées*	-	-	-	-	-
Union libre = oui	23	0	1	272	296
Ménages d'une personne	1	12	4 583	4	4 600
Ménages de deux personnes	0	0	1 154	12	1 166
Ménages de trois personnes	2	22	189	47	260
Ménages de quatre personnes	23	145	52	37	257
Ménages de cinq personnes	193	997	865	92	2 147
Ménages de six personnes ou plus *	-	-	-	-	-
Hommes ≥ 15 ans	0	1	136	3	140
Personnes ≥ 15 ans	0	0	1	0	1
Total des ménages	0	0	0	0	0

* Indique que la caractéristique n'a pas été utilisée comme contrainte parce qu'il s'agissait d'une contrainte redondante.

4. Évaluation des divers niveaux géographiques

Une étude a été faite pour comparer les différences absolues entre les estimations-échantillon et les chiffres de population pour 31 caractéristiques observées en 1996 (29 de celles-ci étaient des contraintes) et en 1986, à différents niveaux géographiques. Les 31 caractéristiques incluses dans cette étude des différences absolues sont présentées à l'annexe B. Les résultats de l'étude sont résumés au tableau 6.6 ci-dessous. Le tableau indique le pourcentage de caractéristiques ayant une « valeur R » se situant à l'intérieur d'un certain intervalle pour les six niveaux géographiques qui figurent dans le tableau. Une valeur R est un quotient des différences de 1996 et celles de 1986 calculé selon l'expression suivante :

contraînte; par conséquent, l'un des sous-groupes doit être redondant), le fait de la retrancher n'aura aucune incidence sur les estimations finales. On appelle contraîntes **faibles**, les contraîntes qui s'appliquent à moins de 20 ménages, tandis que les contraîntes **quasi linéairement dépendantes (QLD)** et les **observations aberrantes** correspondent à des coefficients de pondération supérieurs à 25 ou inférieurs à 0.

Le tableau montre le nombre de fois où les contraîntes ont été retranchées. Ainsi, pour un total de 5 941 RP, la contraînte « Âge 0 à 4 » a été retranchée 3 154 fois, ce qui laisse supposer qu'il existe un risque d'écart entre le chiffre de population et l'estimation-échantillon obtenu à l'aide de coefficients de pondération par régression. L'écart associé à cette caractéristique est très faible (-208 personnes de 0 à 4 ans, selon le tableau 6.4), car il s'agit d'une contraînte **linéairement dépendante**. Cependant, en ce qui concerne les ménages d'une personne, la différence augmente, car cette contraînte a été retranchée parce qu'il s'agit d'une contraînte **quasi linéairement dépendante**. Il convient de noter que, dans le cas des ménages d'une personne, la contraînte a été retranchée 4 600 fois (dans 4 583 cas, parce qu'il s'agissait d'une contraînte QLD) et que, par conséquent, l'écart entre le chiffre de population et l'estimation obtenue au moyen de coefficients de pondération par régression est de -4 740. Il existe un lien évident entre le nombre de fois où une contraînte est retranchée et la différence qui en résulte, pourvu que la contraînte ne soit pas retranchée en raison de sa dépendance linéaire.

Il existe un groupe de caractéristiques dont les résultats sont différents. Soulignons que les caractéristiques « Âge 0 à 4 », « Âge 5 à 9 » et « Âge 10 à 14 » ont toutes été retranchées à une fréquence plutôt élevée. Cependant, les contraîntes « Personnes de plus de 15 ans » et « Population totale » sont presque toujours retenues. Ainsi, l'important groupe d'âge de 0 à 14 ans constitue une contraînte redondante, ce qui explique pourquoi les trois groupes d'âge sont retranchés avec une telle régularité en raison de leur dépendance linéaire.

Tableau 6.4 Écarts estimation/population à l'échelle du Canada en 1996

Contrainte	Coefficients de pondération initiaux	Premier exercice		Deuxième exercice	
		Écart	Différence	Écart	Différence
Hommes	-22 868	-	0,00	15	0,00
Hommes de plus de 14 ans	-48 269	213	0,00	-276	0,00
Personnes de plus de 14 ans	-29 285	3	0,00	3	0,00
Total des ménages	1 060	-	0,00	-	0,00
Population totale	23 117	-	0,00	-	0,00
Âge 0 à 4	15 779	34	0,00	-208	-0,01
Âge 5 à 9	18 705	168	0,01	-258	-0,01
Âge 10 à 14	17 918	-205	-0,01	462	0,02
Âge 15 à 19	4 709	1 890	0,10	1 853	0,10
Âge 20 à 24	-24 353	1 501	0,08	803	0,04
Âge 25 à 29	-17 831	-49	-0,00	105	0,01
Âge 30 à 34	-3 979	248	0,01	361	0,02
Âge 35 à 39	-3 924	493	0,02	320	0,01
Âge 40 à 44	5 251	232	0,01	366	0,02
Âge 45 à 49	9 004	1 743	0,09	971	0,05
Âge 50 à 54	8 267	959	0,06	993	0,06
Âge 55 à 59	-2 135	-201	-0,02	254	0,02
Âge 60 à 64	2 533	-3 380	-0,29	3 847	0,33
Âge 65 à 74 #	4 582	-1 056	-0,05	-662	-0,03
Âge 75 et plus	-11 408	-2 377	-0,19	-9 207	-0,74
Personnes célibataires	-37 340	239	0,00	115	0,00
Personnes mariées	91 338	10	0,00	73	0,00
Veufs et veuves	-11 803	-1 338	-0,10	-1 387	-0,11
Personnes divorcées	-13 606	951	0,06	1 209	0,08
Personnes séparées	-5 472	137	0,02	-10	-0,00
Union libre = oui	-1 404	46 646	2,63	2 415	0,14
Ménages d'une personne #	-	-	-	-4 750	-0,18
Ménages de deux personnes	12 060	-344	-0,01	-3 331	-0,05
Ménages de trois personnes	4 772	3 317	0,06	2 614	0,05
Ménages de quatre personnes	11 666	4 822	0,07	6 776	0,09
Ménages de cinq personnes	3 170	57 775	1,57	27 879	0,76
Ménages de six personnes ou plus #	-	-65 570	-1,38	-29 187	-0,52

= Contrainte ajoutée pour le deuxième exercice.

3. Retranchement des contraintes

Lors du premier exercice, 29 contraintes (voir l'annexe B) au total ont été utilisées au niveau des RP, y compris les groupes d'âge quinquennaux, l'état matrimonial, le sexe et la taille du ménage. La contrainte « État matrimonial – Séparé » n'a pas été utilisée, car il s'agit d'une contrainte linéairement dépendante aux autres contraintes associées à l'état matrimonial. De la même façon, les contraintes « Âge 60 à 64 » et « Ménage de six personnes ou plus » n'ont pas été utilisées parce qu'elles sont linéairement dépendantes à d'autres contraintes. Comme nous l'avons mentionné précédemment, en vue de conserver un plus grand nombre de contraintes, nous avons utilisé en 1996 des coefficients de pondération se situant dans un intervalle de 0 à 25, au lieu d'un intervalle de 1 à 25 comme en 1991.

Le tableau 6.5 présenté ci-après montre la fréquence à laquelle les contraintes ont été retranchées lors du deuxième exercice au niveau des RP et les raisons s'y rattachant. Ainsi, comme une contrainte linéairement dépendante (LD) est redondante (ménages de six personnes ou plus, par exemple, puisque l'ensemble des ménages est déjà une

2. Écart entre les chiffres de population et les estimations-échantillon

Un des buts de la méthode de pondération consiste à réduire les écarts entre les chiffres de population et les estimations-échantillon correspondantes pour les contraintes. Ces écarts sont le résultat de la variabilité et du biais d'échantillonnage (voir le chapitre V). Certains écarts peuvent toutefois subsister même après l'application de la méthode de pondération. Les écarts correspondent à la différence entre les estimations-échantillon et les chiffres de population, laquelle est exprimée en tant que pourcentage des chiffres de population, c'est-à-dire :

$$\text{écart} = \frac{\text{estimation-échantillon} - \text{chiffre de population}}{\text{chiffre de population}} \times 100$$

(2)

Le numérateur de cette expression (estimation-échantillon – chiffre de population) est souvent appelé la « **différence** ». On notera que les chiffres de population correspondent aux SD échantillonnés uniquement. Le tableau 6.4 montre les **différences** et les **écarts** observés à l'échelle du Canada en 1996 pour 32 caractéristiques démographiques de base. On présente les **différences** pour trois estimations-échantillon distinctes, à savoir les coefficients de pondération initiaux, les coefficients de pondération établis lors du premier exercice et les coefficients de pondération établis lors du deuxième exercice. Soulignons que les écarts ont été arrondis à deux décimales. Les estimations-échantillon et les chiffres de population sont fondés sur les logements privés occupés dans les SD échantillonnés.

Le tableau montre clairement que, lors du premier exercice, le système de pondération a permis de réduire les différences pour la plupart des caractéristiques. Ainsi, les coefficients de pondération initiaux ont produit un écart de 91 338 entre les estimations-échantillon et les chiffres de population pour les personnes mariées, tandis que lors du premier exercice de pondération, on a réduit cet écart à 10. Les écarts associés aux groupes d'âge plus jeunes ont aussi diminué considérablement. Le système de pondération n'a toutefois pas permis de réduire l'écart pour toutes les caractéristiques. En fait, pour ce qui est de la caractéristique « union libre », les coefficients de pondération initiaux ont donné une différence de -1 404, que le système de pondération a fait passer à 46 646. Des différences similaires ont également été observées au chapitre de la taille des ménages, en particulier pour ce qui est des ménages de cinq et six personnes. Mentionnons que, dans le cas des ménages de cinq personnes, les coefficients de pondération initiaux ont donné une différence de 3 170, que le système de pondération a fait passer à 57 775. Ces résultats constituent la principale raison pour laquelle l'exercice a été effectué une deuxième fois.

Comme nous l'avons déjà mentionné, lors du deuxième exercice de pondération, la caractéristique « union libre » a été utilisée comme contrainte. En outre, les coefficients de pondération initiaux ont été ajustés de manière à obtenir une concordance avec les chiffres de population associés à la taille des ménages au niveau des RP. Le tableau 6.4 fait également état des différences observées lors du deuxième exercice. Les principales difficultés rencontrées lors du premier exercice ont été résolues lors du deuxième. Ainsi, l'écart associé à la caractéristique « union libre » est tombé de 46 646 à 2 415 et la différence enregistrée pour les ménages de cinq personnes a chuté de 57 775 à 27 879. Cependant, des résultats un peu plus élevés ont été enregistrés pour certaines caractéristiques lors du deuxième exercice. Les personnes âgées de 75 ans et plus, par exemple, affichaient une différence de -2 377 lors du premier exercice, mais cette différence est passée à -9 207 lors du deuxième.

L'évaluation du système de pondération s'est déroulée en quatre étapes. Premièrement, la répartition des coefficients de pondération définitifs a été examinée. Deuxièmement, les écarts entre les chiffres de population et les estimations par régression finales ont été affichés et comparés aux chiffres correspondants de 1991. Troisièmement, les contraintes qui avaient été retranscrites lors du premier exercice ont été décrites brièvement, puis tous les détails associés aux contraintes retranscrites lors du second exercice ont été affichés. Enfin, le rendement du système de pondération a été évalué aux divers niveaux géographiques.

1. Répartition des coefficients de pondération

Le tableau 6.3 montre la répartition des coefficients de pondération pour les deux exercices de production réalisés en 1996 et la répartition correspondante pour 1991. Il est à noter que, en vue de conserver un plus grand nombre de contraintes qu'en 1991, il a été décidé de permettre l'utilisation de coefficients de pondération inclus dans l'intervalle de 0 à 1, de sorte que les intervalles allaient de 0 à 25 en 1996 alors qu'ils allaient de 1 à 25 en 1991.

Tableau 6.3 Répartition des coefficients de pondération de ménages

Intervalles des coefficients de pondération	1991			1996 - Deuxième			1996 - Première		
	Nombre	%		Nombre	%		Nombre	%	
< 1,00	0	-		13 833	0,7		13 611	0,6	
1,00 à 1,99	58 934	3,1		70 195	3,4		68 116	3,3	
2,00 à 2,99	170 352	8,9		204 749	9,8		198 052	9,5	
3,00 à 3,99	338 334	17,7		383 921	18,4		378 736	18,1	
4,00 à 4,99	432 603	22,6		461 281	22,1		465 983	22,3	
5,00 à 5,99	369 274	19,3		379 204	18,1		388 137	18,6	
6,00 à 6,99	242 000	12,7		245 652	11,8		250,521	12,0	
7,00 à 7,99	137 472	7,2		143 887	6,9		144 662	6,9	
8,00 à 8,99	74 270	3,9		80 913	3,9		80 486	3,9	
9,00 à 9,99	39 671	2,1		44 793	2,1		43 947	2,1	
10,00 à 14,99	45 808	2,4		56 569	2,7		53 348	2,6	
15,00 à 19,99	3 182	0,2		4 660	0,2		4 099	0,2	
20,00 à 24,99	357	0,0		525	0,0		484	0,0	
≥ 25,00	0	-		0	-		0	-	
Total	1 912 257	100,0		2 090 182	100,0		2 090 182	100,0	

Le tableau révèle que seule une petite proportion de ménages, soit moins de 0,7 %, avaient un coefficient de pondération inférieur à 1. Dans l'ensemble, on observe une plus forte proportion de coefficients de pondération élevés ou faibles en 1996 qu'en 1991. Mentionnons qu'en 1996 tous les intervalles inférieurs à 3 comptaient pour un pourcentage plus important du total que le groupe correspondant en 1991. En outre, tous les intervalles de 9 ou plus équivalaient également à un pourcentage plus élevé que le groupe correspondant en 1991. Cependant, pour ce qui est des intervalles intermédiaires, entre 4 et 9, l'échantillon de 1991 renfermait un pourcentage plus élevé de ménages que celui de 1996. Si on examine le pourcentage de ménages correspondant aux intervalles situés entre 3 et 7 en 1986 (Bankier et coll., 1992) et en 1991 par rapport à 1996, on constate une diminution marquée. En 1986, alors que l'on utilisait la méthode itérative du quotient, on trouvait 93,9 % des ménages dans ces intervalles; en 1991, ce pourcentage était tombé à 72,3 %, puis il est descendu encore en 1996 pour se chiffrer à 71 % (premier exercice) et à 70,3 % (deuxième exercice).

Tableau 6.2 Nombre de SDR, SR et CEF qui respectent les limites de RP

Description	Limites géographiques de 1996		
	SDR	SR	CEF
Régions géographiques renfermant seulement une RP complète ou une partie d'une seule RP	5 373 (89,8 %)	4 396 (73,5 %)	0 (0 %)
Régions géographiques renfermant plus d'une RP complète	329 (5,5 %)	313 (5,2 %)	46 (15,6 %)
Régions géographiques chevauchant au moins une limite de RP	282 (4,7 %)	1 272 (21,3 %)	249 (84,4 %)
Total	5 984	5 980	295

Le tableau indique que 282 SDR (4,7 %), 1 272 SR (21,3 %) et 249 CEF (84,4 %) ne respectent pas les limites des RP.

Pour obtenir de plus amples renseignements au sujet des régions de pondération et de leurs limites, consulter Kruszynski, 1999.

B. Évaluation de la Procédure d'estimation généralisée des moindres carrés

Le système de pondération de 1991 a été utilisé en 1996 et le logiciel n'a été que très légèrement modifié. Bien que cela suppose que l'on a apporté peu de changement à l'ensemble du système, il n'en reste pas moins que les paramètres (les contraintes définies, par exemple) pouvaient être modifiés. L'objectif était de conserver une proportion plus élevée de contraintes qu'on ne l'avait fait en 1991. En outre, en raison de restrictions budgétaires, certaines des caractéristiques 2A utilisées en 1991 ont été éliminées.

Durant le premier exercice de production, les contraintes associées à la taille des ménages ont souvent été abandonnées au niveau des RP, ce qui s'est traduit par une différence entre les chiffres de population et les estimations pour ces contraintes. On a observé un écart important de 2,6 % pour la caractéristique « Union libre = oui » (qui n'avait pas été utilisée comme contrainte lors du premier exercice). Ces observations donnent à penser que le système de pondération n'a pas réussi à éliminer complètement les biais dans l'échantillon du recensement et que les coefficients de régression pourraient alors donner lieu à des estimations biaisées en ce qui concerne les caractéristiques 2B. C'est pourquoi il a été décidé de mettre le système de pondération à l'essai une deuxième fois de la façon décrite dans le chapitre III, section D du présent rapport.

Pour réduire l'écart associé à la caractéristique « Union libre = oui », celle-ci a été ajoutée comme contrainte lors du deuxième exercice. Par ailleurs, en vue de réduire les écarts observés quant à la taille des ménages, on a ajusté les coefficients de pondération initiaux de manière à obtenir une concordance avec les chiffres de population enregistrés pour les diverses tailles de ménages au niveau des RP. Lors du premier exercice, les coefficients de pondération initiaux pour tous les ménages d'une RP donnée correspondaient simplement au nombre de ménages de la RP divisé par le nombre de ménages échantillonnés dans la RP : ces coefficients de pondération sont appelés **coefficients de pondération initiaux simples**. Lors du second exercice, les coefficients de pondération initiaux simples ont été ajustés pour concorder avec le nombre estimatif de ménages – pour chacune des six tailles de ménage (1, 2, 3, 4, 5, 6+) – au niveau des RP. Ainsi, au départ, avant de mettre à l'essai le système de pondération par régression, une concordance a été établie au niveau des RP entre les estimations-échantillon et les chiffres de population pour les six tailles de ménage. Ces coefficients de pondération ajustés sont appelés **coefficients de pondération initiaux stratifiés a posteriori**. Le système de pondération par régression a ensuite été appliqué à ces coefficients de pondération. Pour obtenir de plus amples renseignements, consulter Bankier, Houle et Luc, 1997.

VI. Évaluation des méthodes de pondération

A. Formation des régions de pondération (RP)

La première étape du processus de pondération a été la formation des RP. Une RP est la plus petite unité géographique pour laquelle une concordance des caractéristiques de la population entre certains résultats-échantillon et chiffres de population peut être assurée. La formation de RP plus grandes permet une meilleure cohérence entre les chiffres de population et les estimations pour un plus grand nombre de caractéristiques démographiques, mais au détriment de la cohérence entre les chiffres de population et les estimations pour les petites régions. En conséquence, en ce qui concerne la taille des RP (nombre de ménages), nous devons faire un compromis entre la nécessité de faire concorder les estimations-échantillon et les chiffres de population des petites régions, d'une part, et la nécessité de faire en sorte que cette concordance s'applique à un vaste éventail de caractéristiques. Pour le recensement de 1996, la formation des RP, qui consistait à regrouper des secteurs de dénombrement (SD), devait respecter les conditions suivantes :

- Une RP doit respecter les limites des divisions de recensement (DR).
 - Une RP devrait contenir entre 800 et 3 000 ménages.
 - Une RP devrait, dans la mesure du possible, respecter les limites de certaines régions de niveau supérieur, la priorité étant donnée, par ordre d'importance, aux subdivisions de recensement (SDR), aux secteurs de recensement (SR) et, enfin, aux circonscriptions électorales fédérales (CEF).
 - Les RP devraient être formées de SD contigus (c.-à-d. être continues) et être le plus compacte possible.
- Étant donné que la Procédure d'estimation généralisée des moindres carrés (PEGMC) est effectuée de façon indépendante dans chaque RP, on obtiendra une concordance entre les estimations-échantillon et les chiffres de population uniquement pour les régions géographiques qui renferment exclusivement des RP complètes. La concordance n'est pas assurée pour les régions géographiques faisant intégralement partie d'une portion de RP ou qui contiennent des portions de différentes RP.
- Le tableau suivant montre la répartition des RP selon leur taille (nombre de ménages).

Tableau 6.1 Répartition des régions de pondération de 1996 selon le nombre de logements et la contiguïté

Logements		RP
0 à 999	4	
1 000 à 1 499	1 686	
1 500 à 1 999	2 213	
2 000 à 2 499	1 417	
2 500 à 2 999	560	
3 000+	61	
Total	5 941	

Le tableau 6.1 montre que 5 876 RP (98,9 %) se situaient à l'intérieur de l'intervalle souhaité (1 000 à 3 000 ménages) en 1996. Par ailleurs, 5 888 RP (99,1 %) étaient contiguës. La plupart des autres RP n'étaient pas contiguës en raison de la non-contiguïté d'un SD ou d'une SDR.

Le tableau suivant montre le nombre de régions géographiques dont les limites respectent celles des RP.

Tableau 5.4 Comparaison entre 1991 et 1996

Caractéristique		1996		1991	
	T-Test	Moyenne	% DR > 0	T-Test	Moyenne
Hommes	-1,1	-0,1	47	5,1	0,3
Femmes	8,5	0,6	70	12,0	0,8
Population totale	5,0	0,3	60	11,2	0,7
Âge 0 à 4	5,3	0,3	60	5,2	0,3
Âge 5 à 9	5,3	0,3	61	7,6	0,5
Âge 10 à 14	4,9	0,3	60	8,2	0,5
Âge 15 à 19	1,7	0,1	55	3,4	0,2
Âge 20 à 24	-6,2	-0,4	37	-3,6	-0,2
Âge 25 à 29	-2,0	-0,1	48	-2,5	-0,2
Âge 30 à 34	2,4	0,2	60	3,2	0,2
Âge 35 à 39	0,6	0,0	54	3,0	0,2
Âge 40 à 44	2,1	0,1	53	4,3	0,3
Âge 45 à 49	3,3	0,2	59	4,3	0,3
Âge 50 à 54	2,1	0,1	56	2,3	0,1
Âge 55 à 59	-0,8	-0,1	50	-0,1	0,0
Âge 60 à 64	0,8	0,1	53	1,0	0,1
Âge 65 à 74	-0,4	0,0	51	-0,2	-0,0
Âge 75 et plus	-7,1	-0,4	31	-6,6	-0,4
Personnes célibataires	-1,9	-0,1	47	3,2	0,2
Personnes mariées	14,2	0,9	82	17,4	1,1
Veufs et veuves	-7,1	-0,4	32	-6,1	-0,4
Personnes divorcées	-3,8	-0,2	43	-5,2	-0,3
Personnes séparées	-3,7	-0,2	40	-5,2	-0,3
Ménages d'une personne	-11,4	-0,7	26	-17,1	-1,0
Ménages de deux personnes	3,2	0,2	56	3,1	0,2
Ménages de trois personnes	2,5	0,2	55	3,6	0,2
Ménages de quatre personnes	4,4	0,3	62	7,6	0,5
Ménages de cinq personnes	2,7	0,2	57	4,6	0,3
Ménages de six personnes ou plus	-4,2	-0,3	41	-1,4	-0,1

Tableau 5.3 Valeurs Z aux niveaux national et régional

Caractéristique					
Hommes	-3,98	2,81	0,33	2,10	-2,91
Femmes	8,28	2,81	3,87	7,57	4,29
Population totale	2,81	2,81	2,34	6,94	0,89
Âge 0 à 4	5,14	5,14	0,33	4,37	2,91
Âge 5 à 9	6,00	5,73	3,75	4,67	3,26
Âge 10 à 14	5,73	1,61	1,03	2,88	3,81
Âge 15 à 19	1,53	1,03	2,03	0,24	0,15
Âge 20 à 24	-8,08	-2,46	-2,35	-4,78	-5,97
Âge 25 à 29	-5,84	-0,81	-0,26	-6,31	-2,84
Âge 30 à 34	-1,20	2,15	1,78	-3,51	-0,92
Âge 35 à 39	-1,18	0,36	-0,52	-0,98	-0,77
Âge 40 à 44	1,65	1,30	2,10	-0,79	1,30
Âge 45 à 49	2,96	1,47	2,15	1,22	1,31
Âge 50 à 54	3,05	-0,52	2,29	2,19	1,29
Âge 55 à 59	-0,87	1,41	-0,33	-1,96	0,20
Âge 60 à 64	1,08	0,11	3,02	-1,00	0,26
Âge 65 à 74	1,49	-0,68	-1,02	2,93	0,70
Âge 75 et plus	-4,89	-3,54	-2,73	-1,46	-2,95
Personnes célibataires	-5,10	-0,58	0,87	-6,55	-2,69
Personnes mariées	-15,03	6,10	9,16	7,12	8,03
Veufs et veuves	-5,54	-2,66	-2,01	-2,98	-3,61
Personnes divorcées	-5,21	-1,98	-0,98	-3,60	-3,85
Personnes séparées	-3,27	-1,51	-0,30	-2,28	-2,26
Union libre = oui	-0,44	-0,54	1,09	-1,23	-0,74
Hommes ≥ 15	-10,85	-2,31	-1,27	-9,23	-7,19
Femmes ≥ 15	4,74	3,04	6,27	-0,32	1,75
Ménages d'une personne	-10,42	-5,02	-8,38	-3,20	-5,01
Ménages de deux personnes	4,00	1,70	0,68	3,32	2,17
Ménages de trois personnes	1,96	0,15	2,79	0,55	0,26
Ménages de quatre personnes	4,91	2,76	4,32	1,17	2,29
Ménages de cinq personnes	1,93	-0,02	2,43	-0,21	1,71
Ménages de six personnes ou plus	-5,71	-0,74	-0,80	-5,18	-3,12

Tableau 5.2 Statistiques sommaires des valeurs Z au niveau des divisions de recensement (DR)

Caractéristique	Moyenne	Type ¹	T ²	% DR (Z > 2)	% DR (Z < -2)	% DR (Z > 0)
Hommes	-0,07	1,08	-1,10	2,1	3,6	47,3
Femmes	0,55	1,08	8,50	8,2	0,4	69,7
Population totale	0,32	1,06	4,98	6,0	0,7	59,8
Âge 0 à 4	0,33	1,06	5,26	6,4	2,5	60,1
Âge 5 à 9	0,33	1,04	5,29	4,3	0,7	61,2
Âge 10 à 14	0,29	0,99	4,87	3,9	0,7	60,1
Âge 15 à 19	0,10	0,98	1,73	3,6	0,7	54,8
Âge 20 à 24	-0,38	1,04	-6,20	1,1	7,1	37,4
Âge 25 à 29	-0,13	1,07	-2,03	1,8	2,8	48,4
Âge 30 à 34	0,15	1,02	2,38	3,2	2,8	60,1
Âge 35 à 39	0,03	0,98	0,60	2,1	1,8	53,7
Âge 40 à 44	0,13	1,01	2,10	3,6	0,7	53,4
Âge 45 à 49	0,22	1,10	3,29	5,3	2,1	59,1
Âge 50 à 54	0,13	1,03	2,12	4,3	2,1	55,5
Âge 55 à 59	-0,05	0,99	-0,78	1,8	2,1	49,8
Âge 60 à 64	0,05	1,10	0,82	2,5	4,6	53,4
Âge 65 à 74	-0,02	1,03	-0,40	1,8	2,1	50,5
Âge 75 et plus	-0,43	1,01	-7,11	0,4	7,8	31,0
Personnes célibataires	-0,13	1,10	-1,91	1,1	3,9	46,6
Personnes mariées	0,87	1,03	14,15	12,5	0,4	81,9
Veufs et veuves	-0,41	0,96	-7,06	0,4	4,3	31,7
Personnes divorcées	-0,23	1,02	-3,78	0,7	6,0	42,7
Personnes séparées	-0,21	0,94	-3,68	1,4	2,1	39,9
Union libre = oui	-0,02	0,91	-0,33	0,7	1,4	52,0
Hommes ≥ 15	-0,39	1,19	-5,50	1,8	6,8	36,7
Femmes ≥ 15	0,34	1,08	5,22	6,4	1,1	62,3
Ménages d'une personne	-0,69	1,02	-11,37	0,0	9,6	26,3
Ménages de deux personnes	0,20	1,04	3,18	5,3	2,1	55,9
Ménages de trois personnes	0,15	0,98	2,54	3,2	0,7	54,8
Ménages de quatre personnes	0,30	1,12	4,44	3,2	2,5	62,3
Ménages de cinq personnes	0,16	0,96	2,74	1,1	1,4	56,9
Ménages de six personnes ou plus	-0,26	1,03	-4,18	1,4	5,3	40,9

¹ Type : déviation type des valeurs Z

² T : statistique étudiante pour tester que moyenne = 0

Tableau 5.1 Statistiques sommaires au niveau du Canada

Caractéristique	Chiffres	Estimation ¹	Différence ²	Écart ³	E.T. ⁴	Valeur Z ⁵
Hommes	13 717 654	13 694 786	(22 868)	-0,17	5 752	-3,98
Femmes	14 176 680	14 222 665	45 985	0,32	5 552	8,28
Population totale	27 894 334	27 917 451	23 117	0,08	8 227	2,81
Âge 0 à 4	1 858 332	1 874 111	15 779	0,85	3 073	5,14
Âge 5 à 9	1 932 023	1 950 728	18 705	0,97	3 120	6,00
Âge 10 à 14	1 939 776	1 957 694	17 918	0,92	3 125	5,73
Âge 15 à 19	1 903 023	1 907 732	4 709	0,25	3 074	1,53
Âge 20 à 24	1 840 654	1 816 301	(24 353)	-1,32	3 013	-8,08
Âge 25 à 29	1 971 123	1 953 292	(17 831)	-0,90	3 053	-5,84
Âge 30 à 34	2 405 559	2 401 580	(3 979)	-0,17	3 317	-1,20
Âge 35 à 39	2 486 060	2 482 136	(3 924)	-0,16	3 339	-1,18
Âge 40 à 44	2 268 423	2 273 674	5 251	0,23	3 177	1,65
Âge 45 à 49	2 050 229	2 059 233	9 004	0,44	3 040	2,96
Âge 50 à 54	1 581 484	1 589 751	8 267	0,52	2 707	3,05
Âge 55 à 59	1 271 221	1 269 086	(2 135)	-0,17	2 448	-0,87
Âge 60 à 64	1 157 926	1 160 459	2 533	0,22	2 338	1,08
Âge 65 à 74	1 991 721	1 996 303	4 582	0,23	3 068	1,49
Âge 75 et plus	1 236 780	1 225 372	-11 408	-0,92	2 332	-4,89
Personnes célibataires	12 779 218	12 741 878	(37 340)	-0,29	7 320	-5,10
Personnes mariées	11 537 475	11 628 813	91 338	0,79	6 076	15,03
Veufs et veuves	1 303 304	1 291 501	(11 803)	-0,91	2 130	-5,54
Personnes divorcées	1 605 136	1 591 530	-13 606	-0,85	2 612	-5,21
Personnes séparées	669 201	663 729	-5 472	-0,82	1 675	-3,27
Union libre = oui	1 770 338	1 768 774	-1 564	-0,09	3 568	-0,44
Hommes ≥ 15	10 781 073	10 732 804	-48 269	-0,45	4 449	-10,85
Femmes ≥ 15	11 383 130	11 402 113	18 983	0,17	4 006	4,74
Ménages d'une personne	2 584 348	2 558 041	-26 307	-1,02	2 524	-10,42
Ménages de deux personnes	3 385 597	3 397 657	12 060	0,36	3 011	4,00
Ménages de trois personnes	1 804 304	1 809 076	4 772	0,26	2 435	1,96
Ménages de quatre personnes	1 813 493	1 825 159	11 666	0,64	2 378	4,91
Ménages de cinq personnes	737 751	740 921	3 170	0,43	1 640	1,93
Ménages de six personnes ou plus	334 207	327 786	-6 421	-1,92	1 124	-5,71

¹ Fondée sur les coefficients de pondération simples

² Différence : estimation-chiffre

³ Écart : écart (100*[différence-chiffre]/chiffre)

⁴ E.T. : erreur-type de l'estimation du coefficient de pondération simple

⁵ Valeur Z : (estimation-biais-chiffre)/E.T.

retranchés de l'échantillon (c'est-à-dire qu'ils ont été convertis de ménages 2B en ménages 2A pour devenir des ménages non inclus dans l'échantillon) et seules les réponses aux questions de base ont été imputées. Ce processus de conversion des ménages inclus dans l'échantillon en ménages non inclus dans l'échantillon est appelé conversion de document 2A/2B. Un petit nombre (52) de ménages 2A et 4 ont été convertis en ménages 2B en vue de satisfaire à certaines contraintes de pondération. Il est possible que les ménages manquants ou refusant de répondre, ainsi que les ménages n'ayant pas répondu aux questions-échantillon, aient des caractéristiques différentes des autres ménages (ils pourraient être, par exemple, de plus petite taille). Ainsi, la conversion de ménages 2B en ménages 2A pourrait biaiser l'échantillon. Par ailleurs, si le système d'imputation avait tendance à imputer plus souvent certaines caractéristiques aux ménages manquants ou refusant de répondre qu'à d'autres types de ménages, des divergences entre les estimations-échantillon et les chiffres de population auraient été produites, car seuls des ménages non inclus dans l'échantillon seraient touchés.

La variation géographique du biais a aussi été étudiée. Le tableau 5.3 montre les valeurs Z correspondant aux 32 caractéristiques pour l'ensemble du Canada, la région de l'est, le Québec, l'Ontario et la région de l'ouest. Bien qu'il soit difficile de tirer des conclusions définitives de ce tableau, étant donné que la taille des échantillons varie énormément, certaines tendances sont évidentes. Le Québec se distingue des autres régions en ce qui concerne les hommes. En effet, ceux-ci sont surreprésentés dans l'échantillon du Québec, alors qu'ils sont sous-représentés dans le reste du pays. Cependant, l'Ontario se démarque également des autres régions pour ce qui est de la population totale. Le nombre de célibataires et de personnes vivant en union libre est aussi surreprésenté au Québec, tandis qu'il est sous-représenté dans le reste du Canada. Mentionnons que ce résultat reflète peut-être davantage une différence dans la façon dont l'état matrimonial est enregistré au Québec. Il est à noter qu'il y a concordance quant aux groupes d'âge d'une région à l'autre, pour ce qui est de l'orientation du biais, exception faite de quelques différences mineures pour les personnes d'âge moyen. Les personnes âgées de plus de 15 ans sont représentées à peu près de la même façon dans toutes les régions, les hommes étant sous-représentés et les femmes surreprésentés (sauf en Ontario, où les femmes sont très légèrement sous-représentées). Enfin, le biais suit également une tendance assez constante en ce qui touche la taille des ménages. Dans l'ensemble, malgré les quelques différences observées entre les régions, on observe de nombreuses similitudes.

Enfin, le tableau 5.4 présente brièvement quelques résultats comparatifs entre l'Examen du biais d'échantillonnage de 1991 et celui de 1996. La comparaison des résultats au niveau des DR révèle que l'orientation du biais n'a changé que dans le cas des hommes et des célibataires. En 1991, ces catégories de personnes étaient surreprésentées dans l'échantillon, alors qu'elles étaient sous-représentées en 1996. Si l'on compare le nombre de DR affichant des valeurs Z supérieures à 0, on constate une différence tout à fait saisissante entre les deux échantillons en ce qui touche les hommes. En 1991, les hommes étaient surreprésentés dans 63 % des DR, tandis que cette proportion était de 47 % en 1996. Le nombre de DR dans lesquelles les célibataires étaient sous-représentés constitue l'autre changement important : en 1991, les célibataires étaient surreprésentés dans 56 % des DR, comparativement à 47 % en 1996. Cependant, au chapitre des groupes d'âge, les deux échantillons présentaient des similitudes remarquables; seules les personnes âgées de 35 à 44 ans étaient surreprésentées dans un plus grand pourcentage de DR en 1991 par rapport à 1996.

En résumé, il existe des biais dans l'échantillon 2B et ces biais sont semblables à ceux qui ont été observés dans l'échantillon du recensement de 1991. Nous ne pouvons toutefois pas déterminer à quoi ces biais sont attribuables ni ce qu'il conviendrait de faire pour les éliminer. Cependant, la méthode de pondération choisie (la PEGMC) contribuera dans une large mesure à corriger les biais pour les caractéristiques choisies comme contraintes. Reste à savoir quels sont les biais que l'on observera pour les caractéristiques n'ayant pas été retenues comme contraintes.

Cette formule est expliquée plus en détail à l'annexe C. La statistique obtenue a été calculée à l'aide de données comportant des imputations pour chacune des 281 divisions de recensement (DR) du Canada et pour l'ensemble des 32 caractéristiques où $X^{(0)}$ est une estimation fondée sur l'échantillon 2B établie au moyen de coefficients de pondération simples, tandis que X est le chiffre de population connu (2A) et que $V(X^{(0)})$ représente la variance d'échantillonnage de l'estimateur $X^{(0)}$. Les valeurs $Z^{(0)}$, pour les 281 DR, devraient suivre une distribution à peu près normale (moyenne = 0; variance = 1) si un échantillon aléatoire simple – ou un échantillon systématique – de ménages était sélectionné de façon non biaisée dans chacun des SD et si il n'était pas touché par le traitement.

Le tableau 5.1 présente les résultats à l'échelle nationale, mais il ne permet pas de déterminer si quelques résultats présentent un biais négatif compensant d'autres résultats présentant un biais positif. Pour trouver réponse à cette question, on présente au tableau 5.2 un résumé des statistiques relatives au biais pour les 281 DR. Le tableau montre la moyenne, l'erreur-type et le résultat du test effectué (qui consistait à déterminer si la moyenne réelle des valeurs Z était de 0) pour les valeurs Z . Cette hypothèse était rejetée si $|T|$ des valeurs Z était inférieur à 1,96 pour les 32 caractéristiques de base. En règle générale, les caractéristiques pour lesquelles des résultats biaisés avaient été enregistrés (tableau 5.1) à l'échelle nationale donnaient également des résultats biaisés si l'on examinait la totalité des DR. Toutefois, certaines différences sont dignes de mention. Ainsi, au niveau national, les hommes semblaient sous-représentés dans l'échantillon, mais ce biais n'a pas été observé lors de l'examen des DR. Comme, on **présume**, dans le tableau 5.2, **que toutes les DR sont de la même taille**, nous pouvons en déduire que l'estimation des hommes dans les grandes DR (urbaines) présente un biais négatif, tandis que les hommes des petites DR (rurales) sont probablement surreprésentés. En revanche, un biais a été observé à l'échelle nationale en ce qui concerne les ménages de cinq personnes, mais l'examen des DR a révélé qu'ils étaient surreprésentés dans l'échantillon. Ce résultat s'explique par le fait que, même si très peu de DR présentent une sous-représentation ou une surreprésentation élevée, on observe une faible, mais constante, surreprésentation de ce groupe dans tout l'échantillon.

Le tableau 5.2 montre également le pourcentage de DR correspondant des valeurs Z inférieures à -2, supérieures à 2 et supérieures à 0. Si l'échantillon n'était pas biaisé, nous nous attendrions, d'une part, à ce qu'environ 2,5 % des DR affichent des valeurs Z inférieures à -2 ou supérieures à 2 et, d'autre part, à ce que 50 % des DR correspondent à des valeurs Z supérieures à 0. De toute évidence, certaines caractéristiques prouvent que le biais est répandu parmi presque toutes les DR. Par exemple, les personnes mariées sont surreprésentées dans 82 % des DR. Par ailleurs, dans 13 % des DR, les valeurs Z associées aux personnes mariées sont supérieures à 2, alors qu'elles sont inférieures à -2 dans seulement 0,4 % des DR, soit 1 sur 281. L'échantillon du recensement contient manifestement un trop grand nombre de personnes mariées. Par contre, le nombre de ménages d'une personne était surreprésenté (valeur Z supérieure à 0) dans seulement 26 % des DR. Mentionnons également qu'aucune DR n'a obtenu une valeur Z supérieure à 2 pour les ménages d'une personne.

Comme on l'a mentionné dans l'introduction, les différences observées entre les estimations-échantillon (basées sur une pondération simple) et les chiffres de population peuvent être expliquées de différentes façons. Une cause possible est le fait qu'au recensement de 1996, il n'y a eu que 86 183 ménages (0,8 % du total) qui ont été manquants ou qui ont refusé de répondre aux questions, en comparaison avec 253 156 (2,6 % du total) au recensement de 1991. Ces ménages ont totalement refusé de répondre aux questions ou le recenseur a été incapable d'obtenir des renseignements parce que les membres du ménage étaient absents au moment où le recensement a été mené ou avaient déménagé, le jour même du recensement ou après cette date, sans répondre au questionnaire. Le recenseur pouvait parfois déterminer le nombre de personnes, mais en général toutes les autres réponses devaient être imputées pour ces ménages. De l'ensemble des ménages manquants ou ayant refusé de répondre en 1996, 32 820 (38 % des 86 183 ménages) étaient des ménages de l'échantillon, alors que nous serions attendus à ce que 20 % en fassent partie si les taux de non-réponse des ménages 2A et 2B étaient équivalents. En outre, 3 358 ménages de l'échantillon, même s'ils n'étaient pas manquants ou n'avaient pas refusé de répondre (c'est-à-dire qu'ils avaient fourni certaines réponses aux questions de base), n'ont fourni aucune réponse aux questions-échantillon. Au cours du traitement des données, ces 38 820 + 3 358 = 42 179 ménages de l'échantillon n'ayant fourni aucune réponse aux questions-échantillon ont été

V. Biais d'échantillonnage

Les estimations établies à partir d'une enquête-échantillon sont sujettes à des erreurs d'échantillonnage. Un type d'erreur d'échantillonnage est causé par la variabilité des caractéristiques de la population. Cette variabilité provient d'échantillons différents qui produisent des estimations différentes, aucune d'elles n'égalant nécessairement la vraie valeur de la population. Toutefois, les estimations égaleront la vraie valeur de la population en moyenne **pourvu que l'échantillon ne comporte pas de biais**. La présence de biais donne lieu à une tendance à la surestimation ou à la sous-estimation de la vraie valeur de la population. Malheureusement, il est souvent difficile d'éliminer le biais complètement. Dans le cas du recensement de la population, les réponses peuvent être entachées d'un biais causé, entre autres, par les erreurs de couverture, le biais de non-réponse (p. ex., des recensés qui répondent différemment sur la formule 2B que sur la formule 2A), les erreurs commises par les recenseurs (p. ex., des recenseurs ne choisissant pas l'échantillon selon les exigences établies), les erreurs de traitement, et ainsi de suite.

Le recensement de la population constitue une occasion unique d'examiner avec attention le biais associé aux caractéristiques de base. Il existe essentiellement deux estimations des mêmes caractéristiques de base : les chiffres pour la population totale et les estimations-échantillon (1 sur 5). Le but de l'Examen du biais d'échantillonnage était de déterminer si les réponses aux questions de base de la formule 2B comportaient un biais. On a obtenu les estimations-échantillon en multipliant les chiffres de l'échantillon au niveau du SD par des coefficients de pondération simples égaux à la fraction de sondage inverse des ménages du SD (environ 5) et en faisant ensuite la somme au niveau de la région géographique appropriée⁵.

Au départ, les résultats ont été calculés à l'échelle nationale. Le tableau 5.1 montre les résultats obtenus au niveau national pour les chiffres de population et les estimations ainsi que la différence entre ces résultats, l'écart et l'erreur-type des estimations établies pour 32 caractéristiques de base. La valeur Z (qui constitue un test de signification statistique) figure également au tableau; ceux-ci sont expliqués plus en détail ci-dessous. Dans le cas de 22 caractéristiques, la valeur Z se situe à l'extérieur des intervalles prévus, à savoir -2 à 2. Ainsi, les valeurs Z associées aux hommes, aux personnes de 20 à 29 ans, aux personnes de 75 ans et plus et aux hommes de plus de 15 ans sont inférieures à -2, ce qui indique que ces groupes sont sous-représentés dans l'échantillon, au niveau national, par rapport à l'ensemble de la population. Cependant, les valeurs Z associées aux femmes, à la population totale, aux personnes de 0 à 14 ans et aux personnes mariées sont supérieures à 2, ce qui révèle que ces caractéristiques sont surreprésentées dans l'échantillon. En outre, les ménages d'une personne et les ménages de six personnes ou plus étaient sous-représentés, tandis que les ménages de 2 à 5 personnes étaient surreprésentés. En général, l'échantillon contient un trop grand nombre d'enfants, de personnes d'âge moyen, de femmes et de personnes mariées, tandis qu'il ne renferme pas suffisamment de jeunes adultes, de personnes âgées, d'hommes et de personnes non mariées.

Ces différences, qui étaient statistiquement significatives au seuil de confiance de 5 % dans le tableau 5.1, ont été établies à l'aide de la formule suivante :

$$Z^{(0)} = \frac{\bar{X}^{(0)} - X}{\sqrt{V(\bar{X}^{(0)})}} \quad (1)$$

⁵ Ces estimations simples ont été utilisées au lieu des estimations PEGMC, parce que la PEGMC tend à masquer le biais d'échantillonnage en forçant les estimations de caractéristiques de base à égaler les chiffres de population.

D. Variance d'échantillonnage

La « variance » d'une estimation est une mesure de sa précision. Il est relativement peu coûteux de calculer des estimations de la variance pour des estimateurs lorsqu'on utilise des coefficients de pondération égaux à 5 et qu'on suppose un échantillonnage aléatoire simple. Toutefois, il est beaucoup plus coûteux d'estimer la variance des estimateurs du recensement pour le plan d'échantillonnage et les techniques d'estimation utilisés. Des « **facteurs d'ajustement** » ont été calculés pour le recensement de 1986. Les facteurs d'ajustement sont le rapport entre les estimations des erreurs-types (les racines carrées de la variance) des estimations du recensement et les estimations simples des erreurs-types. On peut alors obtenir une estimation de l'erreur-type d'une estimation du recensement pour n'importe quelle caractéristique dans toute unité géographique en multipliant l'estimation simple de l'erreur-type par le facteur d'ajustement approprié. On explique pourquoi ces estimations de l'erreur-type ne sont pas nécessairement exactes, en raison du biais introduit dans le processus par l'échantillonnage, le traitement des données et la méthode d'estimation.

IV. Programme d'évaluation de l'échantillonnage et de la pondération

Le programme d'évaluation de l'échantillonnage et de la pondération avait pour objet de déterminer l'effet de l'échantillonnage et de la pondération sur la qualité des données-échantillon du recensement. À cette fin, quatre études ont été réalisées afin de mesurer la qualité des données-échantillon et des estimations-échantillon du recensement ainsi que de fournir des renseignements utiles à la planification des futurs recensements. Les études ont porté sur les sujets suivants :

- a) examen du biais d'échantillonnage;
- b) évaluation des méthodes de pondération;
- c) évaluation de la cohérence des estimations-échantillon et des chiffres de population;
- d) étude visant à évaluer la variance d'échantillonnage pour diverses caractéristiques-échantillon (20 %).

La suite du présent chapitre décrit brièvement ces quatre études. Les chapitres V à VIII en présentent les résultats.

A. Examen du biais d'échantillonnage

Dans toute enquête, un biais entachant les réponses peut provenir de diverses sources. Le but de cette étude était de déterminer si les réponses aux questions de base du questionnaire 2B étaient biaisées de quelque façon que ce soit et de cerner, si possible, les causes de tout biais observé.

B. Évaluation des méthodes de pondération

Le but de cette étude était d'évaluer l'efficacité de la PEGMC. À cette fin, on a examiné le niveau de concordance entre les estimations-échantillon et les chiffres de population à l'égard des diverses contraintes, pour l'ensemble des RP au Canada. Le nombre et le type des contraintes retranchées au niveau des RP, ainsi que les raisons pour lesquelles elles avaient été retranchées, ont été étudiés pour expliquer les incohérences observées. En outre, on a examiné la distribution des coefficients de pondération PEGMC, ainsi que les écarts entre les résultats de 1996 et ceux de 1991.

C. Examen de la cohérence des estimations-échantillon et des chiffres de population

Dans cette étude, on a examiné le niveau de concordance, ou de cohérence, entre les estimations-échantillon et les chiffres de population pour un large éventail de caractéristiques de base. On a étudié cette concordance pour diverses unités géographiques autres que les RP. On a également comparé la cohérence obtenue en 1996 à celle obtenue en 1991 pour ces caractéristiques.

totalité des personnes dans les ménages de taille 1 sont célibataires, les deux contraintes sont donc presque équivalentes, et l'on peut en considérer une comme QLD. Les contraintes LD ont été retranchées pour accroître l'efficacité de calcul de l'algorithme de pondération. Les contraintes faibles et les contraintes QLD ont été retranchées parce que, si on ne l'avait pas fait, les estimations auraient pu devenir instables et présenter des erreurs-types élevées.

Une fois retranchées les contraintes faibles, LD et QLD au niveau des RP, le calcul des coefficients de pondération PEGMC a lieu en deux étapes. À la première étape, les coefficients de pondération initiaux, qui égalent l'inverse de la fraction d'échantillonnage des ménages dans les SD, sont ajustés individuellement pour chaque SD. Il faut noter que certaines contraintes, qui n'ont pas été retranchées au niveau des RP, peuvent l'être au niveau des SD, parce qu'elles sont faibles ou linéairement dépendantes. Les contraintes qui demeurent pour le SD sont triées selon le nombre de ménages auxquels elles s'appliquent au niveau du SD. Elles sont ensuite divisées en deux groupes, soit ceux de nombre pair et ceux de nombre impair. Les coefficients de pondération PEGMC sont calculés au niveau du SD pour chaque groupe de contraintes. Parfois, la méthode d'estimation donne des coefficients de pondération très faibles (inférieurs à 10) ou très élevés (supérieur à 25) pour obtenir la concordance nécessaire à l'égard de certaines contraintes. Ces coefficients, appelés coefficients de « valeur aberrante », ne sont pas désirables. Par conséquent, quand cela se produit, les contraintes qui en sont responsables sont repérées et retranchées, et les coefficients sont calculés de nouveau. Enfin, on fait la moyenne des coefficients obtenus pour les deux groupes de contraintes pour chaque ménage de l'échantillon, ce qui donne les coefficients de pondération de la première étape pour chaque SD.

Les coefficients de pondération produits à la première étape sont utilisés comme coefficients initiaux de la deuxième étape, qui consiste à les ajuster de façon à obtenir une concordance entre les estimations-échantillon et les chiffres de population au niveau des RP. Toutes les contraintes non désignées comme faibles, LD or QLD au niveau de la RP sont utilisées. Ici encore, si des valeurs aberrantes sont produites, les contraintes responsables sont repérées et retranchées, et les coefficients définitifs sont calculés de nouveau. Même si la deuxième étape détruit quelque peu la concordance obtenue pour les estimations au niveau des SD à la première étape, les estimations finales au niveau des SD sont tout de même plus proches des chiffres de population que si la première étape avait été omise. Pour une explication plus détaillée du calcul des coefficients de pondération, voir Bankier, Rathwell et Majkowski, 1992.

En 1996, deux estimations distinctes ont été effectuées. La deuxième estimation a été jugée nécessaire lorsque certaines divergences inacceptables ont été observées pour les variables « union libre » et « taille du ménage » lors de la première estimation. Pour la deuxième estimation, une nouvelle contrainte (union libre = oui) a été ajoutée et le coefficient de pondération initial a été modifié. Lors de la première estimation, le coefficient de pondération initial des ménages correspondait à l'inverse de la fraction d'échantillonnage des ménages. Cependant, ces coefficients de pondération initiaux ont été ajustés lors de la deuxième estimation pour obtenir une concordance entre les estimations relatives à la taille des ménages et les chiffres de population au niveau des RP. Pour obtenir de plus amples renseignements, consulter Bankier, Houle et Luc, 1997.

Des coefficients de pondération ont été établis à l'aide de la PEGMC uniquement pour les ménages privés ayant reçu le questionnaire complet du recensement parmi un SD échantillonné (1 sur 5). Un coefficient de pondération de 0 a été attribué aux ménages privés du SD échantillonné qui avaient reçu une formule abrégée. On a attribué un coefficient de pondération 1 à tous les ménages privés des SD non échantillonnés, car la totalité des répondants de ces secteurs ont fourni des renseignements sur la formule 2B. Les ménages collectifs ont aussi reçu un coefficient de pondération de 1.

Par exemple, si l'on recensait 5 000 hommes de 20 à 24 ans à l'Île-du-Prince-Édouard, et si 1 020 de ces hommes faisaient partie des ménages inclus dans l'échantillon, un coefficient de pondération de $5\,000/1\,020 = 4,90$ serait attribué à chaque homme de 20 à 24 ans dans l'échantillon de l'Île-du-Prince-Édouard. De cette façon, chaque fois que les groupes d'âge-sexe de cinq ans seraient recoupés avec une caractéristique-échantillon pour l'Île-du-Prince-Édouard, le total marginal pour le groupe d'âge-sexe d'hommes de 20 à 24 ans concorderait avec le total de la population, soit 5 000. Ce genre de méthode d'estimation est appelée « estimation par la méthode du quotient ». Il est à noter que, dans cet exemple particulier, un coefficient de pondération de 5 donnerait une estimation-échantillon de 5 100 (1 020 x 5). La méthode d'estimation qui a été utilisée avant le recensement de 1991 était une généralisation de l'estimation par quotient appelée méthode itérative du quotient (MIQ). Pour plus de détails sur la MIQ, voir le *Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération* (Statistique Canada, 1990), ainsi que Brackstone et Rao, 1979.

Pour le recensement de 1991, on a décidé de recourir à une autre méthode d'estimation, appelée procédure d'estimation généralisée des moindres carrés (PEGMC) en deux étapes. Ce choix visait à améliorer la concordance entre les chiffres de population et les estimations correspondantes au niveau des SD, par rapport à ce que permettait la MIQ. Les erreurs-types des estimations produites par la PEGMC pour les petites régions géographiques étaient également réduites. En outre, la PEGMC permettait de déterminer, pour chaque ménage de l'échantillon, un coefficient de pondération unique pouvant servir à produire des estimations tant pour les caractéristiques des personnes que pour les caractéristiques des ménages. Par contre, avec la MIQ, il fallait utiliser des coefficients de pondération différents pour produire des estimations des caractéristiques des ménages et des personnes, ce qui entraînait parfois des incohérences.

Avec la PEGMC, les coefficients de pondération d'environ 5 ont été ajustés aussi peu que possible pour chaque ménage, de sorte qu'on s'assurait qu'il y avait une concordance parfaite entre les estimations et les chiffres de population pour autant de caractéristiques de base que possible qui sont énoncées à l'annexe B (celles-ci seront appelées contraintes). Il fallait que cette concordance parfaite soit réalisée au niveau des régions de pondération (RP). Chaque RP contenait, en moyenne, sept SD échantillonnés. On retrouve plus d'information sur les RP au chapitre VI, section A du présent rapport. On y explique que la PEGMC est un estimateur par régression. L'ouvrage de Sarnadal, Swensson et Wretman, 1992, qui renferme plus de détails sur les estimateurs par régression, constitue un bon outil de référence.

D. Procédure d'estimation généralisée des moindres carrés (PEGMC) en deux étapes

Les calculs des coefficients de pondération sont faits indépendamment dans chaque région de pondération (RP). Certaines des contraintes énoncées à l'annexe B (tant au niveau des SD qu'à celui des RP) doivent être retranchées pour chaque RP, de sorte que la concordance entre les chiffres de population et les estimations ne peut être garantie pour toutes les contraintes. Les contraintes sont initialement retranchées au niveau des RP pour les raisons suivantes :

- elles s'appliquent à moins de 20 ménages (on les appelle des contraintes **faibles**);
- elles sont redondantes (on les appelle des contraintes **linéairement dépendantes** [LD]);
- elles sont quasi redondantes (on les appelle des contraintes **quasi linéairement dépendantes** [QLD]).

Une contrainte redondante serait, par exemple, le nombre total de femmes, puisque les contraintes sont déjà présentes lorsque le nombre total de personnes et le nombre total d'hommes sont utilisés. On est en présence d'une **contrainte linéairement dépendante** lorsque deux contraintes (ou plus) font en sorte que la troisième sera automatiquement satisfaite. Pour illustrer ce qu'on entend par **contrainte quasi redondante**, prenons le cas des contraintes représentant les personnes dont l'état matrimonial = célibataire et taille du ménage = 1. Si la plupart, mais non la

Heureusement, ces deux objectifs sont généralement complémentaires, c'est-à-dire que l'erreur d'échantillonnage est habituellement réduite lorsqu'on s'assure que les estimations-échantillon de certaines caractéristiques de base sont compatibles avec les chiffres correspondants de la population. Notons cependant qu'il peut arriver que le fait de forcer les estimations-échantillon de caractéristiques de base à être compatibles avec les chiffres correspondants de la population pour de très petits sous-groupes puisse avoir des conséquences défavorables sur l'erreur-type des estimations des caractéristiques-échantillon proprement dites.

Lorsqu'on n'a aucune information sur la population échantillonnée autre que celle qui a été recueillie pour les unités d'échantillonnage, la méthode d'estimation se limite à pondérer les unités d'échantillonnage en proportion inverse de leur probabilité de sélection; par exemple, si toutes les unités avaient une chance sur cinq d'être choisies, toutes les unités recevraient un coefficient de pondération de 5. En pratique, toutefois, on a généralement une certaine connaissance de la population observée, par exemple, sa taille globale et, peut-être, sa répartition en fonction d'une variable donnée (disons, par province). Ces données peuvent être utilisées pour améliorer la formule d'estimation de façon à produire des estimations qui se rapprochent davantage de la valeur inconnue de la population. Dans le cas de l'échantillon du recensement, on dispose d'un grand nombre de données très détaillées sur la population échantillonnée, soit les caractéristiques de base obtenues pour tous les ménages, à tous les niveaux géographiques. On peut se servir de cette multitude d'informations sur la population pour améliorer les estimations faites à partir de l'échantillon du recensement mais, d'autre part, ces données peuvent également occasionner des ennuis compte tenu du fait qu'il est impossible d'obtenir des estimations-échantillon des caractéristiques de base qui soient compatibles avec toutes les données sur la population à tous les niveaux géographiques. Les écarts entre les estimations-échantillon et les valeurs de la population apparaissent quand on produit un recoupement d'une variable de base et d'une variable-échantillon. La totalisation doit être fondée sur les données-échantillon; de ce fait, les totaux marginaux des variables de base sont des estimations-échantillon que l'on peut comparer avec les chiffres correspondants de la population d'une autre totalisation basée sur les données intégrales. Il se peut que cette comparaison ne donne pas une concordance exacte.

C. Elaboration d'une méthode d'estimation pour l'échantillon du recensement

Étant donné qu'il faut attribuer un coefficient de pondération à chaque unité d'échantillonnage (personne, famille ou ménage), le plus simple serait de choisir le coefficient 5, puisqu'il s'agit d'un échantillon de 1 sur 5. Une telle méthode serait simple et sans biais⁴, et si l'on n'avait pas d'autres données que celles de l'échantillon, cela pourrait être la meilleure méthode. Or, bien qu'on sache que l'échantillon contiendra presque exactement le cinquième de tous les ménages (à l'exclusion des ménages collectifs et de ceux qui se trouvent dans les secteurs de recensement par interview), on ne peut pas être certain qu'il contiendra exactement un cinquième de toutes les personnes, ou un cinquième de chaque genre de ménage, ou un cinquième de toutes les femmes de 25 à 34 ans, etc. Par conséquent, une telle méthode n'assurerait pas la cohérence, même pour les sous-groupes les plus importants de la population. Dans ce dernier cas, les fractions seraient très proches d'un cinquième, mais elles pourraient s'en écarter de façon sensible dans le cas des petits sous-groupes. Une autre méthode simple serait de prendre certains sous-groupes importants, par exemple les groupes d'âge-sexe par province, puis, pour chacun, compter le nombre d'éléments de la population appartenant au sous-groupe (N) et le nombre d'éléments appartenant à l'échantillon (n), et attribuer à chaque unité d'échantillonnage appartenant au sous-groupe un coefficient de pondération égal à N/n .

⁴ « Sans biais » veut dire que la moyenne des estimations de l'ensemble de tous les échantillons possibles serait égale à la valeur vraie de la population.

III. Estimations basées sur l'échantillon du recensement

Tout échantillonnage nécessite l'application d'une méthode d'estimation pour rapporter les données-échantillon à l'échelle de la population totale. La méthode d'estimation est généralement choisie en fonction d'exigences d'ordre opérationnel et théorique. Sur le plan opérationnel, cette méthode doit être compatible avec le système de dépouillement auquel elle s'intègre, tandis que sur le plan théorique, elle doit minimiser l'erreur d'échantillonnage des estimations produites. Les deux sections qui suivent décrivent les considérations opérationnelles et théoriques qui régissent le choix des méthodes d'estimation à partir de l'échantillon du recensement.

A. Considérations d'ordre opérationnel

Une méthode d'estimation peut être décrite mathématiquement à l'aide d'une formule algébrique qui indique comment la valeur de l'estimateur de la population est calculée en fonction des valeurs observées dans l'échantillon. Quand il s'agit de petites enquêtes qui ne portent que sur une ou deux caractéristiques, ou lorsque la formule d'estimation est très simple, il est possible de calculer les estimations-échantillon en appliquant cette formule aux données-échantillon pour chaque estimation requise. Toutefois, dans le cas d'une enquête ou d'un recensement portant sur une gamme étendue de caractéristiques, ou lorsque la formule d'estimation est assez complexe, il n'est pas possible d'appliquer une formule séparément pour chaque estimation désirée. Dans le cas du recensement, par exemple, chaque cellule de chaque totalisation, fondée sur des données-échantillon recueillies à chaque niveau géographique, représente une estimation-échantillon qui nécessiterait, sous cette approche, une application distincte de la formule d'estimation. Par ailleurs, les diverses estimations effectuées par calcul individuel à partir d'un même échantillon du recensement ne sont pas nécessairement cohérentes.

Par conséquent, la méthode utilisée pour le recensement et pour de nombreuses enquêtes-échantillon consiste à décomposer le processus d'estimation en deux étapes : a) le calcul des coefficients de pondération, c'est-à-dire la pondération elle-même et b) la sommation des coefficients de pondération afin de produire des estimations des chiffres de population. Les seules difficultés mathématiques, s'il y en a, se rapportent à l'étape a) qui n'est effectuée qu'une fois, tandis que l'étape b) se limite à un simple processus de sommation des coefficients de pondération au moment où une totalisation est extraite. Il faut noter que, comme le coefficient de pondération relié à une unité de l'échantillon est le même quelle que soit la totalisation effectuée, la cohérence entre les diverses estimations basées sur les données-échantillon est assurée.

B. Considérations d'ordre théorique

Selon la théorie de l'échantillonnage, il est possible de déterminer, à partir d'un plan d'échantillonnage particulier et d'une procédure d'estimation donnée, les chances qu'un certain intervalle contienne la valeur (inconnue) que l'on veut estimer pour la population entière. Le principal objet de l'estimation est de réduire au maximum l'étendue de ces intervalles de façon que les hypothèses sur les valeurs inconnues de la population soient aussi précises que possible. La mesure courante de précision utilisée pour comparer les procédures d'estimation est appelée l'erreur-type. Pourvu que certaines conditions relativement peu contraignantes soient respectées, un intervalle de plus ou moins deux erreurs-types par rapport à la valeur estimée renfermera la valeur de la population pour environ 95 % de tous les échantillons possibles.

Outre la réduction au minimum de l'erreur-type, le deuxième objectif relatif au choix d'une méthode d'estimation de l'échantillon du recensement est de garantir autant que possible que les estimations faites à partir de l'échantillon des caractéristiques de base (c.-à-d. 2A) soient cohérentes avec les valeurs connues de la population correspondante.

Dans la terminologie de l'échantillonnage, l'échantillon du recensement peut être décrit comme un échantillon stratifié avec sélection systématique de logements privés occupés, choisi en fonction d'un taux d'échantillonnage constant de 1 sur 5 dans toutes les strates (SD). En tant qu'échantillon de personnes, il peut être considéré comme un échantillon stratifié avec sélection systématique de grappes, un logement constituant une grappe. Sarnadal, Swensson et Wretman, 1992, donnent une description détaillée des concepts et de la terminologie se rapportant à l'échantillonnage.

C. Traitement des données de l'échantillon du recensement

Dès que le recenseur était en possession du questionnaire rempli (2A ou 2B) de chaque logement de son secteur et que son travail avait été approuvé, les questionnaires étaient envoyés à l'un des sept centres de dépouillement régionaux aux fins du traitement manuel. Les données complètes pour chaque SD ont été saisies et stockées sur bandes magnétiques. Les questionnaires et les bandes magnétiques ont ensuite été envoyés au bureau central à Ottawa. Après réception au bureau central, des contrôles automatisés étaient effectués aux fins d'incohérences dans les données, ces incohérences devant être résolues manuellement. Une fois terminées toutes les mises à jour des données pour un SD, les données étaient restructurées et transférées à la section Contrôle et Imputation.

Les données étaient chargées dans les bases de données de contrôle et de l'imputation, organisées par données 2A (100 %) et données 2B (20 %), en cinq régions pour chacune des bases. Les bases de données 2A contenaient les caractéristiques démographiques de base pour l'ensemble de la population et les bases de données 2B contenaient les données-échantillon (20 %). Les données ont été soumises à une série de modules de traitement spécialisés, où tous les problèmes liés aux données invalides, incohérentes ou manquantes ont été résolus. Les bases de données 2A ont été traitées en premier et ont donné lieu à la création de la Base de données d'extraction 2A définitive pour le Canada. Une fois la dernière main mise aux données intégrales, les données-échantillon (20 %) ont été traitées. Les enregistrements de non-réponse 2B ont été supprimés de la base de données 2B. Une base de données d'extraction 2B définitive a été créée. Cette base de données contenait à la fois les données intégrales et les données-échantillon uniquement pour les ménages et les personnes de l'échantillon. Les coefficients de pondération calculés à l'aide des données intégrales (selon la méthode décrite au chapitre III) ont été incluses dans cette base de données.

B. Plan d'échantillonnage utilisé lors du recensement de 1996

Le jour du recensement en 1996, une multitude de renseignements ont été recueillis auprès de chaque Canadien. Il faut noter qu'en 1996, le jour du recensement a été déplacé de la date traditionnelle du début juin pour être fixée au 14 mai 1996. La majeure partie de ces informations ont été obtenues à partir d'un échantillon. Dans chaque secteur d'autodénombrement, un échantillon de 1 sur 5 des logements privés occupés a été choisi pour recevoir un questionnaire complet (formule 2B) contenant toutes les questions du recensement. Les questions de base sur l'âge, le sexe, l'état matrimonial, la langue maternelle et le lien avec la personne repère du ménage (Personne 1) ont été posées à toutes les personnes alors que les questions additionnelles portant sur le genre de logement, le mode d'occupation et les questions socioéconomiques ont été posées à un échantillon de 20 % de la population.

Tous les logements se trouvant dans les secteurs dénombrés par un recenseur (généralement les régions éloignées ou les réserves indiennes) ont reçu le questionnaire 2B. Tous les logements collectifs ont également reçu la formule 2B. Par contre, les personnes suivantes dans les logements collectifs n'ont pas eu à répondre aux questions posées à l'échantillon de la population :

- a) les détenus dans les établissements de correction et les établissements pénitentiaires ou les prisons;
- b) les patients dans les hôpitaux généraux, les établissements de soins spéciaux, les établissements pour les personnes âgées ainsi que les établissements pour personnes souffrant de maladies chroniques et les hôpitaux psychiatriques;
- c) les foyers collectifs pour enfants (orphelins) et les foyers pour jeunes contrevenants.

Les Canadiens en poste à l'étranger (en général, du personnel d'ambassade ou des membres des Forces armées) ont reçu une formule 2C, qui contenait les mêmes questions que la formule 2B mais qui ne comportait pas de questions sur le logement. Les répondants devaient toutefois indiquer leur lieu habituel de résidence au Canada. Les renseignements sur les logements privés inoccupés ont été consignés sur une formule 2A.

D'après les procédures de livraison de base, le recenseur devait planifier un itinéraire couvrant tous les logements de son secteur de dénombrement (SD) puis se rendre à chaque logement et y laisser un questionnaire. Le choix de l'échantillon, c'est-à-dire la décision quant au genre de questionnaire à livrer à chaque logement occupé, a été fait à partir du Registre des visites (RV) dans lequel le recenseur inscrivait tous les logements de son secteur. Chaque cinquième ligne du registre était ombrée, indiquant qu'un questionnaire 2B devait être laissé. Un point de départ aléatoire était déterminé à partir de la première ligne du RV, et le recenseur sautait zéro, une, deux, trois ou quatre lignes selon que le cinquième, quatrième, troisième, deuxième ou premier logement du SD devait être le premier à recevoir le questionnaire complet. Par la suite, chaque logement inscrit sur une ligne ombrée recevait automatiquement le questionnaire complet. Ces procédures étaient expliquées dans le manuel du recenseur et une attention particulière y était portée pendant la formation du recenseur afin de réduire au minimum le risque d'erreur dans la sélection de l'échantillon. Des vérifications du travail des recenseurs, à des fins de contrôle qualitatif, ont été effectuées par le commissaire au recensement.

Pour le recensement de 1996, un essai important sur les opérations du recensement a été réalisé par la poste auprès de plus de 400 000 ménages dans l'est de l'Ontario, y compris la région d'Ottawa. Tous les ménages des régions urbaines ont été inscrits dans un registre des adresses puis des questionnaires ont été expédiés à tous ces logements. Au préalable, les ménages-échantillon (2B) avaient été sélectionnés systématiquement à partir de ce registre. Les répondants étaient invités à retourner leur questionnaire par la poste; un petit groupe de personnes très qualifiées ont assuré un suivi auprès des non-répondants.

sur quelques caractéristiques de base, auprès d'un échantillon d'un tiers de la population au recensement de 1971. Sauf dans les régions les plus isolées du Canada, le tiers des ménages privés a reçu le questionnaire abrégé comportant toutes les questions du recensement, tandis que le reste des ménages privés a reçu le questionnaire complet. Les ménages vivant dans des régions éloignées prédesignées et dans des logements collectifs³ ont reçu le questionnaire complet. Le rapport statistique « Sampling in the Census » (Bureau fédéral de la statistique, 1968) donne une description détaillée des projets d'application de cette méthode au recensement de 1971.

Le contenu du recensement de 1976 était beaucoup moindre que celui de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté de collecte, par exemple les questions sur le revenu, ni celles pour lesquelles les frais de codage sont le plus élevés, par exemple celles sur la profession, l'activité économique et le lieu de travail. Par conséquent, les avantages de l'échantillonnage en termes de coûts et d'allègement du fardeau du répondant étaient moins évidents qu'au recensement de 1971. Néanmoins, après avoir évalué les économies éventuelles reliées à divers taux d'échantillonnage et étudié les conséquences sur le plan des relations publiques du retour au dénombrement intégral (l'échantillonnage ayant fait ses preuves en 1971), il a été décidé d'appliquer en 1976 la même méthode d'échantillonnage qu'en 1971.

Au recensement de 1981, on a utilisé la plupart des mêmes méthodes qu'en 1971 et 1976, sauf que le taux d'échantillonnage est passé du tiers au cinquième des ménages. Les études réalisées à cette époque ont révélé que la diminution consécutive de la qualité des données (mesurée au moyen de la variance) serait acceptable et qu'elle ne serait pas suffisamment importante pour annuler les avantages découlant de la réduction des coûts et du fardeau de réponse et diminuerait les délais de production (voir Royce, 1983). Douze questions ont été posées à l'ensemble de la population et 34 questions de plus ont été posées aux membres de l'échantillon.

Le recensement de 1986 a été le premier recensement majeur réalisé au milieu d'une décennie. Il avait été décidé que seul un recensement majeur pourrait répondre aux besoins croissants en matière de données locales sur le marché du travail, besoins rendus plus pressants par l'importante récession (1981-1982) survenue depuis le recensement précédent. Toutefois, afin de maintenir les coûts d'élaboration au niveau le plus bas possible, on a décidé de faire très peu de changements. Par conséquent, les questions ainsi que les méthodes de collecte et de dépouillement des données du recensement de 1981 ont été reprises, sauf s'il y avait des raisons convaincantes de ne pas procéder ainsi. Des questions portant sur huit sujets posées au recensement de 1981 n'ont pas été reprises en 1986, tandis que trois nouvelles questions ont été ajoutées.

En 1991, le recensement de la population englobait à la fois les résidents permanents et, pour la première fois, les résidents non permanents – personnes qui sont titulaires d'un permis de séjour pour étudiants, d'un permis de travail ou d'un permis ministériel, ou qui revendiquent le statut de réfugié. Afin qu'on puisse savoir qui étaient les résidents non permanents, une nouvelle question a dû être élaborée, testée et incluse dans le recensement de 1991. En tout, 12 nouvelles questions ont été ajoutées au recensement de 1991, tandis que quatre questions du recensement de 1986 n'ont pas été posées en 1991. Une fois terminée la collecte des questionnaires du recensement de 1991, deux enquêtes postcensitaires ont été effectuées. Une de ces enquêtes, l'Enquête sur la santé et les limitations d'activités (ESLA), visait à recueillir des données sur la santé et le bien-être des Canadiens en général. La deuxième enquête, soit celle auprès des peuples autochtones, visait à recueillir des renseignements sur la population autochtone vivant dans les réserves et hors des réserves. De plus, en 1991, on a noté une augmentation importante de l'automatisation du traitement des données et des véhicules par lesquels les produits et services sont élaborés et livrés au client.

³ Un logement collectif est une institution ou un établissement commercial ou communautaire, par exemple un hôtel, un hôpital, une résidence de personnel ou un camp de chantier.

II. Échantillonnage dans les recensements canadiens

Dans un recensement de la population, l'échantillonnage consiste à recueillir et à traiter certaines caractéristiques à partir d'un échantillon aléatoire de logements et de personnes dénombrés lors du recensement intégral. On obtient ensuite pour l'ensemble de la population les totaux pour ces caractéristiques en pondérant les données-échantillon. Les caractéristiques recueillies pour tous les logements ou toutes les personnes faisant partie du recensement sont appelées « caractéristiques de base », tandis que les caractéristiques recueillies à partir d'un échantillon sont appelées « caractéristiques-échantillon » du recensement.

A. Historique de l'échantillonnage au recensement canadien¹

L'échantillonnage a été utilisé pour la première fois au cours du recensement canadien de 1941. Un questionnaire sur le logement a été remis à tous les 10 logements de chaque sous-district de recensement. Les réponses aux 27 questions de ce questionnaire ont été intégrées à celles du questionnaire sur la population et les ménages correspondant à ces logements, ce qui a permis de faire le recoupement des caractéristiques-échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage au cours de l'étape du dépouillement pour obtenir des estimations provisoires du revenu des salariés, de la répartition de la population en âge de travailler et de la composition des familles au Canada. Dans ce cas, l'échantillon était constitué du dixième de tous les secteurs de dénombrement du Canada et tous les questionnaires sur la population de ces secteurs ont été dépouillés à l'avance.

L'échantillonnage a également été utilisé pour le recensement du logement de 1951. Cette fois, on a choisi chaque cinquième logement (ceux dont le numéro d'identification se terminait par 2 ou par 7) et le questionnaire sur le logement comportait 24 questions. Au cours du recensement de 1961, on a prélevé un échantillon de 20 % des ménages privés et on a demandé aux membres de 15 ans et plus de ces ménages de remplir un questionnaire-échantillon sur la population comportant des questions sur la migration interne, la fécondité et le revenu. L'échantillonnage n'a pas été utilisé au cours des recensements de moindre envergure de 1956 et 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de recensement de 1971. La première a été l'utilisation de l'autodénombrement pour la majorité de la population au lieu de la méthode traditionnelle de recensement par représentant. Ce changement découlait des résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964; Hansen et al., 1959) selon lesquels la présence d'un représentant avait un effet important sur la variance² de chiffres du recensement. On prévoyait que l'autodénombrement permettrait de réduire la variance des chiffres du recensement attribuable à la présence du représentant et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire à son aise; par conséquent, on s'attendait à ce que les réponses soient plus exactes que par le passé.

Un autre changement survenu dans le recensement de 1971 a été la modification du contenu. Le nombre de sujets traités et de questions posées était plus élevé qu'aux recensements précédents. Une étude des coûts, du fardeau du répondant et des délais de production par rapport à la qualité des données dans le cadre d'un recensement par échantillonnage et par autodénombrement a présidé à la décision de recueillir toutes les données, sauf celles portant

¹ Des renseignements détaillés sur les recensements particuliers sont contenus dans le Rapport administratif, la Revue générale, le Guide sommaire ou Le recensement en bref des recensements en question. Ces produits sont mentionnés en bibliographie à la fin de ce rapport.

² La « variance » d'une estimation est une mesure de sa précision. La variance est expliquée plus en détail au chapitre VIII.

surdénumbrement qui se sont produits pour diverses raisons; on y a eu recours pour évaluer la qualité des données du recensement. Toutefois, la principale application de l'échantillonnage au recensement a été effectuée pendant le dénombrement sur le terrain où toutes les données, excepté les données de base, ont été recueillies auprès d'un échantillon de ménages. Le présent rapport décrit cette dernière utilisation de l'échantillonnage et présente une évaluation de l'incidence de cette méthode sur la qualité des données du recensement.

Le chapitre II fait l'historique de l'application des méthodes d'échantillonnage aux recensements canadiens et décrit celles qui ont été utilisées au recensement de 1996. Le chapitre III décrit les méthodes de pondération des données-échantillon utilisées en vue d'obtenir des estimations à l'échelle de la population totale et explique les considérations d'ordre pratique et théorique qui régissent le choix de ces méthodes. On présente au chapitre IV les études effectuées dans le cadre du programme d'évaluation des méthodes d'échantillonnage et de pondération utilisées au recensement de 1996. Les chapitres V à VIII donnent les résultats de ces études. Le chapitre IX présente certaines conclusions sur les méthodes de pondération utilisées en 1996 et quelques suggestions pour le recensement de 2001.

I. Introduction

Le recensement de 1996 a nécessité la participation de l'ensemble de la population canadienne, soit environ 29 millions de personnes réparties sur un territoire couvrant 9,2 millions de kilomètres carrés. Une entreprise de cette envergure a constitué un énorme défi à relever. Même si des normes de qualité rigoureuses régissent la collecte et le traitement des données, et malgré les efforts déployés afin de réduire la non-réponse, à l'aide des communications notamment, il est impossible de supprimer toutes les erreurs. Bien que ce terme ne signifie pas nécessairement que des fautes ont été commises, certains éléments d'erreur sont inévitables compte tenu des décisions prises afin de limiter les coûts du recensement.

Statistique Canada doit expliquer les méthodes et les concepts utilisés pour recueillir et traiter les données; il doit en outre renseigner les utilisateurs sur la qualité des données produites, ainsi que sur d'autres caractéristiques des données qui pourraient limiter leur utilité ou leur interprétation. Le présent rapport vise à informer les utilisateurs de la complexité des données et des problèmes susceptibles d'influer sur la façon de les utiliser. Il explique le cadre théorique et les définitions ayant servi à recueillir les données, ainsi que les circonstances inhabituelles pouvant avoir une incidence sur la qualité des données. Par ailleurs, il traite la saisie des données, le contrôle et l'imputation, ainsi que la comparabilité chronologique des données.

Le présent rapport, qui porte sur l'échantillonnage et la pondération, a été préparé par Peter Dick, Karen Switzer, Sylvain Thivierge et Patrick Mason, avec le concours du personnel de deux divisions de Statistique Canada : la Division des méthodes d'enquêtes sociales et la Division des opérations du recensement.

Les utilisateurs trouveront des renseignements supplémentaires sur les concepts, les variables et la géographie du recensement dans le *Dictionnaire du recensement de 1996* (n° 92-351-XPF au catalogue), de même qu'un aperçu des différentes étapes du recensement dans *Le recensement en bref* (n° 92-352-XPF au catalogue).

De nos jours, l'échantillonnage constitue une technique largement utilisée dans maints domaines. Nous pouvons juger de la qualité d'un produit à l'aide d'un échantillon avant d'en faire l'achat; nous formons une opinion des gens à partir d'échantillons de leur comportement; nos impressions sur tel pays ou telle ville sont basées sur les courts voyages que nous y avons faits. Ce sont là des exemples d'échantillonnage où l'on tire des conclusions sur le « tout » à partir de connaissances portant sur une « partie ».

D'une façon plus scientifique, l'échantillonnage est utilisé par les comptables, par exemple, lorsqu'ils effectuent la vérification des états financiers; dans l'industrie, pour vérifier la qualité d'articles produits en série, et par les initiateurs de sondages d'opinion et d'enquêtes pour recueillir des renseignements sur les opinions ou les caractéristiques d'une population donnée. On a généralement recours à l'échantillonnage pour réduire les coûts ou pour obtenir rapidement des données, ou les deux. Dans certains cas, le mesurage peut détruire le produit à l'étude (par exemple, l'évaluation de la durée de vie des ampoules électriques) et par conséquent, l'échantillonnage s'avère essentiel. Le désavantage de cette méthode est que les résultats obtenus à partir d'un échantillon peuvent ne pas être aussi précis que ceux recueillis auprès de la population totale. Toutefois, l'échantillonnage peut être avantageux quand la perte de précision (qui peut être minime si l'échantillon est suffisamment grand) est acceptable en fonction de l'utilisation prévue des résultats. De plus, la réduction de l'envergure d'une étude faite à partir d'un échantillon peut mener à la réduction des erreurs non dues à l'échantillonnage, ce qui compense en partie le manque de précision attribuable à l'échantillonnage.

Pour le recensement de la population de 1996, on a appliqué les méthodes d'échantillonnage de plusieurs façons. On y a eu recours pour s'assurer que la qualité du travail des recenseurs dans le cadre de la collecte des questionnaires respectait certaines normes; on y a eu recours pour le contrôle qualitatif du codage des réponses au cours du dépouillement dans les bureaux régionaux; on y a eu recours pour estimer le sous-dénombrement et le

Annexes	
A. Glossaire	45
B. Contraintes au niveau des RP et des SD appliquées aux coefficients de pondération	45
C. Renseignements additionnels sur les statistiques utilisées dans l'Examen du biais d'échantillonnage	47
D. Facteurs d'ajustement de l'erreur-type de 1986 au niveau national ou provincial et centiles des facteurs au niveau de la région de pondération	51
E. Produits et services	61
Bibliographie	63
Centres régionaux de consultation	65
5.1 Statistiques sommaires au niveau du Canada	16
5.2 Statistiques sommaires des valeurs Z au niveau des divisions de recensement (DR)	17
5.3 Valeurs Z aux niveaux national et régional	18
5.4 Comparaison entre 1991 et 1996	19
6.1 Répartition des régions de pondération de 1996 selon le nombre de logements et la contiguïté	21
6.2 Nombre de SDR, SR et CEF qui respectent les limites de RP	22
6.3 Répartition des coefficients de pondération de ménages	23
6.4 Écarts estimation/population à l'échelle du Canada en 1996	25
6.5 Fréquence à laquelle les contraintes au niveau des RP ont été retranchées en 1996	27
6.6 Pourcentage des caractéristiques ayant des valeurs R dans certains intervalles	28
7.1 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les DR – recensements de 1996 et de 1991	32
7.2 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SDR – recensements de 1996 et de 1991	33
7.3 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SR – recensements de 1996 et de 1991	34
7.4 Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SD – recensements de 1996 et de 1991	35
8.1 Estimations non ajustées des erreurs-types des estimations-échantillon	40

Table des matières

Page

Liste des tableaux	ii
I. Introduction.....	1
II. Échantillonnage dans les recensements canadiens	3
A. Historique de l'échantillonnage au recensement canadien	3
B. Plan d'échantillonnage utilisé lors du recensement de 1996	5
C. Traitement des données de l'échantillon du recensement.....	6
III. Estimations basées sur l'échantillon du recensement.....	7
A. Considérations d'ordre opérationnel.....	7
B. Considérations d'ordre théorique.....	7
C. Elaboration d'une méthode d'estimation pour l'échantillon du recensement	8
D. Procédure d'estimation généralisée des moindres carrés (PEGMC) en deux étapes	9
IV. Programme d'évaluation de l'échantillonnage et de la pondération	11
A. Examen du biais d'échantillonnage	11
B. Évaluation des méthodes de pondération.....	11
C. Examen de la cohérence des estimations-échantillon et des chiffres de population.....	11
D. Variance d'échantillonnage	12
V. Biais d'échantillonnage	13
VI. Évaluation des méthodes de pondération	21
A. Formation des régions de pondération (RP)	21
B. Évaluation de la Procédure d'estimation généralisée des moindres carrés.....	22
1. Répartition des coefficients de pondération	23
2. Écart entre les chiffres de population et les estimations-échantillon	24
3. Retraitement des contraintes	25
4. Évaluation des divers niveaux géographiques	27
VII. Examen de la cohérence entre les estimations-échantillon et les chiffres de population	29
A. Divisions de recensement (DR)	29
B. Subdivisions de recensement (SDR)	30
C. Secteurs de recensement (SR)	30
D. Secteurs de dénombrement (SD)	31
E. Incidence des changements apportés à la méthode de pondération en 1996.....	31
VIII. Variance d'échantillonnage	37
A. Étude sur la variance d'échantillonnage du recensement de 1986	37
B. Variance d'échantillonnage et biais dans l'estimation généralisée des moindres carrés.....	39
IX. Conclusion	41

Données de catalogage avant publication (Canada)

Vedette principale au titre:

Échantillonnage et pondération

(Rapports techniques du recensement de 1996)
Texte en français et en anglais disposé tête-bêche.
Titre de la p. de t. addit.: Sampling and weighting.
ISBN 0-660-60571-6

CS92-371-XPB

1. Échantillonnage (Statistique).
2. Canada – Recensement, 1996 – Méthodologie.
3. Canada – Recensement, 1996 – Évaluation.
1. Statistique Canada. II. Titre. III. Sampling and Weighting. IV. Collection.

HA741.5 1996 S25 1999 304.6'07'23
C99-988039-XF

Pour citer ce document

Statistique Canada. Échantillonnage et pondération.
Rapports techniques du recensement de 1996.
Ottawa : Industrie Canada, 1999.
Numéro 92-371-XPB au catalogue.



Statistique Canada

Échantillonnage et pondération

Rapports techniques du recensement de 1996

Publication autorisée par le ministre responsable de Statistique Canada

© Ministère de l'Industrie, 1999

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division de la commercialisation, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Décembre 1999

N° 92-371-XPB au catalogue

ISBN 0-660-60571-6

Ottawa

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Des données sous plusieurs formes

Statistique Canada diffuse les données sous formes diverses. Outre les publications, des totalisations habituelles et spéciales sont offertes. Les données sont disponibles sur Internet, disque compact, imprimé d'ordinateur, microfilm, et bande magnétique. Des cartes et d'autres documents de référence géographiques sont disponibles pour certaines sortes de données. L'accès direct à des données agrégées est possible par le truchement de CANSIM, la base de données ordinaire et le système d'extraction de Statistique Canada.

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à l'un des centres de consultation régionaux de Statistique Canada suivants :

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(780) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

Vous pouvez également visiter notre site sur le Web : <http://www.statcan.ca>

Un service d'appel interurbain sans frais est offert à tous les utilisateurs qui habitent à l'extérieur des zones de communication locale des centres de consultation régionaux.

Service national de renseignements

Service national d'appareils de télécommunications pour les malentendants
Numéro pour commander seulement (Canada et États-Unis)
Numéro pour commander par télécopieur (Canada et États-Unis)

1 800 263-1136
1 800 363-7629
1 800 267-6677
1 877 287-4369

Renseignements sur les commandes et les abonnements

Les prix ne comprennent pas les taxes de vente

Le produit n° 92-371-XPB au catalogue est offert au coût de 15 \$ par copie au Canada. À l'extérieur du Canada, le coût est de 15 \$ US par copie.

Veillez commander par la poste, en écrivant à Statistique Canada, Division de la diffusion, Gestion de la circulation, 120, avenue Parkdale, Ottawa (Ontario) K1A 0T6; par téléphone, en composant le (613) 951-7277 ou le 1 800 700-1033; par télécopieur, en composant le (613) 951-1584 ou le 1 800 889-9734; ou par Internet, en vous rendant à order@statcan.ca. Lorsque vous signalez un changement d'adresse, veuillez nous fournir l'ancienne et la nouvelle adresses. On peut aussi se procurer les produits de Statistique Canada auprès des agents autorisés, dans les librairies et dans les bureaux régionaux de Statistique Canada.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois et dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec le centre de consultation régional de Statistique Canada le plus près de chez vous.



92-371-XPB96000



N° 92-371-XPB au catalogue

Échantillonnage et pondération

Rapports techniques du recensement de 1996



Statistique
Canada
Statistics
Canada



Canada